

# Measuring the User Experience

David Batten

August 2014

## ***Resources used in this presentation***

- *Practical Usability Testing* - Human Factors International
- *Measuring the User Experience* (2013) - Tullis & Albert
- UXPA Website
- And many more.....

## ***Introduction***

- Me
- 'Advisory' HF Engineer at IBM for 13 years
  - IBM Division representation to the IBM UX Advisory Council
    - 'Tony' Anthony C. Temple, former VP Ease of Use, IBM
  - Worked on lots of products
- Director of Web Usability and Content Compliance - University of Texas for 3 years
  - OMG, what a web page... used as an example later on
- VP User Research & Insights - Citibank for 2 years

## Stuff I've worked on

- IBM POS 4680-4690 Systems



- Restaurants
- Post Offices
- WalMarts

- LCD Touch Screen Angle - [‘International Journal of Industrial Ergonomics’](#) – “Optimal Viewing Angle for Touch-Screen Displays: Is There Such a Thing?” 22, 343-350.
- 4690 MICR Reader Check Printer - IBM Internal Publication
- Placement of POS MCR - IBM Internal Publication
- NCR DynaKey vs IBM SurePoint Touch - IBM Internal Publication
- [Color Coding to Facilitate Performance of Focused Attention Tasks with Object Displays](#). Proceedings of the Human Factors Society 36th Annual Meeting, Vol. 2, 1493 – 1497.

See, Research Can be Published !



## Stuff I've worked on

- Software for IBM Networking Hardware



- 2216 Pre-sales Configurator (web based)
- 2216 Re-Purpose documentation
  - Drop in Replacement? HA!
- Stand alone Configuration Tool for ALL IBM networking hardware (Portfolio of over 50 products)
- Designed NHD's first 'Install-Shield' like PC software installer

- Coca-Cola
- Telecom Italia
- Deutsche Bank
- \$250K box

## Stuff I've worked on

- Configuration Software for IBM p-Series Servers



- AIX Operating System-Web Based System Manager
- Patent Issued (IBM) #US20030217132 "System and method for remotely managing a computer system by a wireless communications device" (2003)
- Cluster Configuration
- Network Visualization for management software



- \$1.4M a single box

...and more stuff

## *Philosophy - Pragmatism*

- Deliver high quality findings within
  - Budget - There's always a cost
  - Time - There's always a time limit
  - Constraints - # of people, equipment
- Have Fun doing it!



Someone's been listening to Devo!

# Basics

Research / UCD / Data Types / Statistics

August 2014



## What is Research?

# re·search

*/ˈrē,sərCH,riˈsərCH/* 

*noun*

1. the systematic investigation into and study of materials and sources in order to establish facts and reach new conclusions.

"we are fighting meningitis by raising money for medical research"

*synonyms:* investigation, experimentation, testing, analysis, fact-finding, fieldwork, examination, scrutiny [More](#)

*verb*

1. investigate systematically.

"she has spent the last five years researching her people's history"

*synonyms:* investigate, study, inquire into, look into, probe, explore, analyze, examine, scrutinize, review [More](#)

*...and the KeyWord is?*

## What is Research?

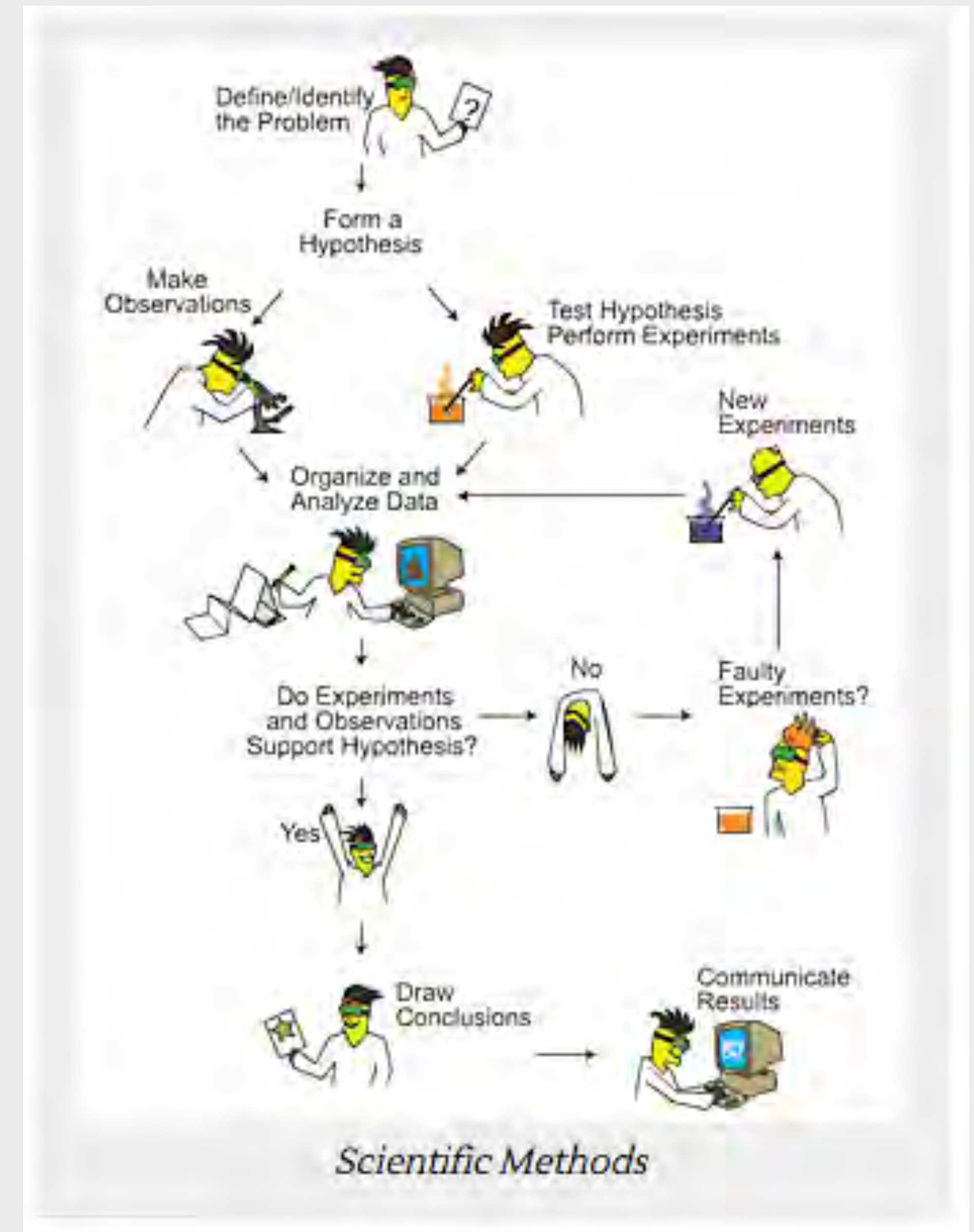
In scientific circles 'Research' is defined as "A Process by which the 'Scientific Method is followed'.

The scientific method is the important process by which all scientific knowledge is acquired. It is a tried and tested method that has been refined over the centuries leading to ever greater discoveries and a better understanding of the universe around us.

The scientific method began with the rules of logic established by the Greek philosopher Aristotle. Over time other philosophers and scientists improved on his work refining the process of inquiry and proving of theories and hypotheses. The current version of the method is 6 to 8 steps depending on whether you are looking to explain an observed phenomenon, coming up with new methods, or integrating old information.

The first step is to define the question. You look at the problem you are trying solve or the phenomena you are trying to understand and formulate a question that can get a solution. This step is the most important as asking the right question is more likely to lead you to the right answer.

The next step is to collect data and observe. This is the part where you either study previous bodies of knowledge or observe the phenomena for the first set of clues needed to find the answer to your question. Observation when done properly will draw your attention to information you may miss at the first glance.



# *What is Research?*

- Define Your Objective (Make it better or seek universal truths?)
- Observation
- Question
- Test
- Accept/Reject hypotheses
- Observation
- More Questions
- Test

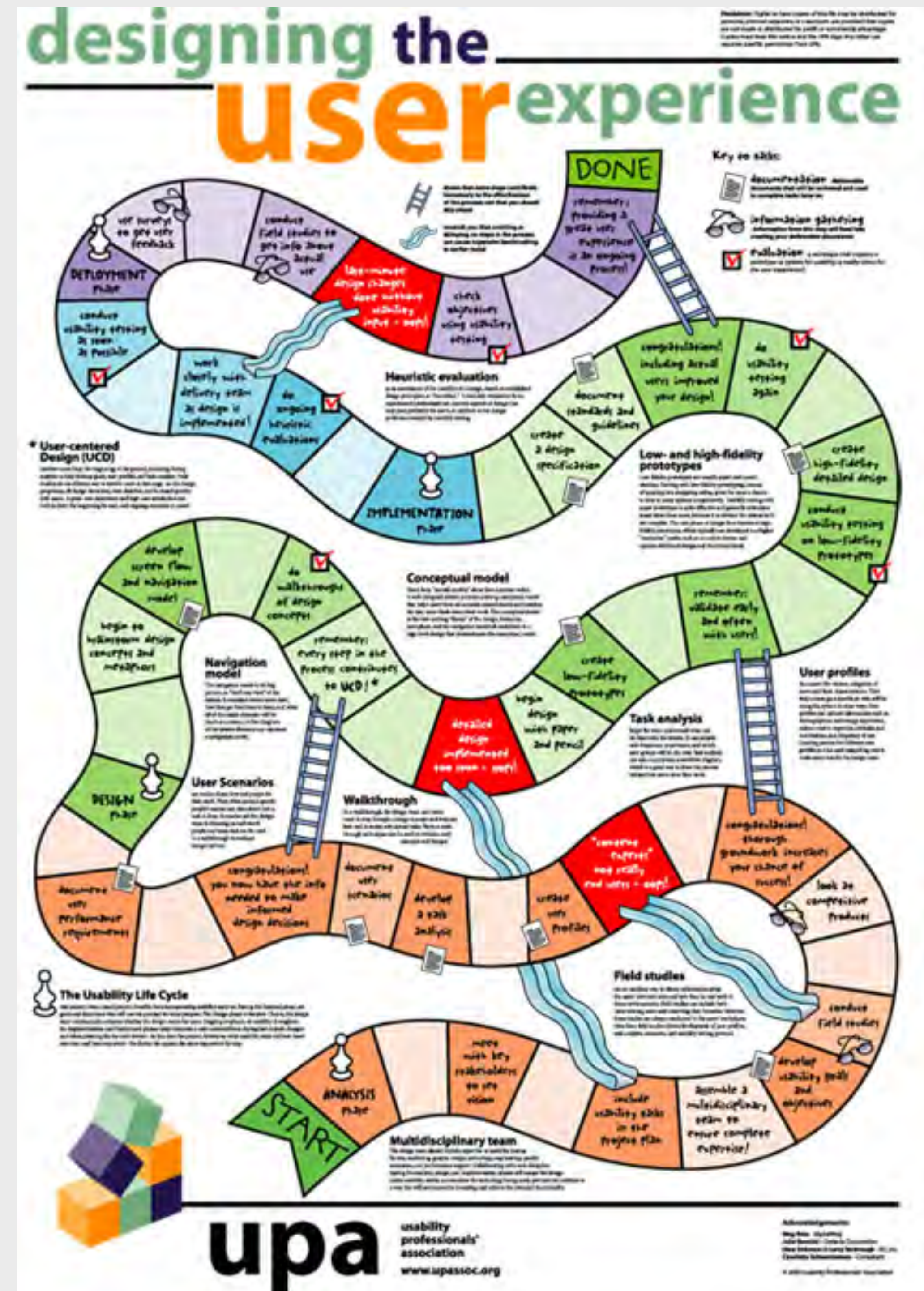
.....

Read more: <http://www.universetoday.com/74036/what-are-the-steps-of-the-scientific-method/#ixzz39Rjm6RMj>

# What is User Centered Design?

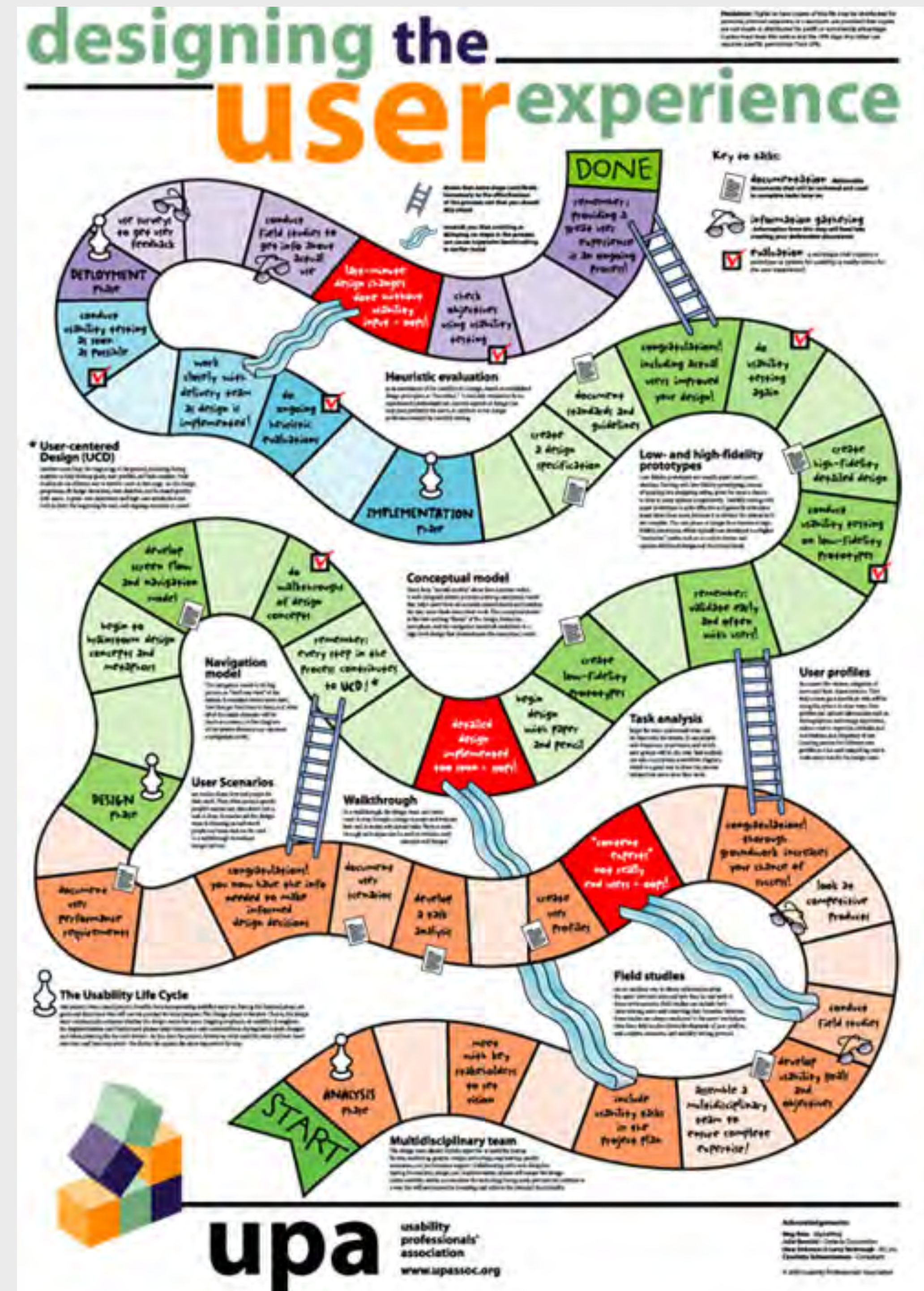
User-centered design (UCD) is an approach to design that grounds the process in information about the people who will use the product. UCD processes focus on users through the planning, design and development of a product.

There is an international standard that is the basis for UCD methodologies.



# What is User Centered Design?

This standard ([ISO 9241: Human-centered design process](#)) defines a general process for including human-centered activities throughout a development life-cycle, but does not specify exact methods.

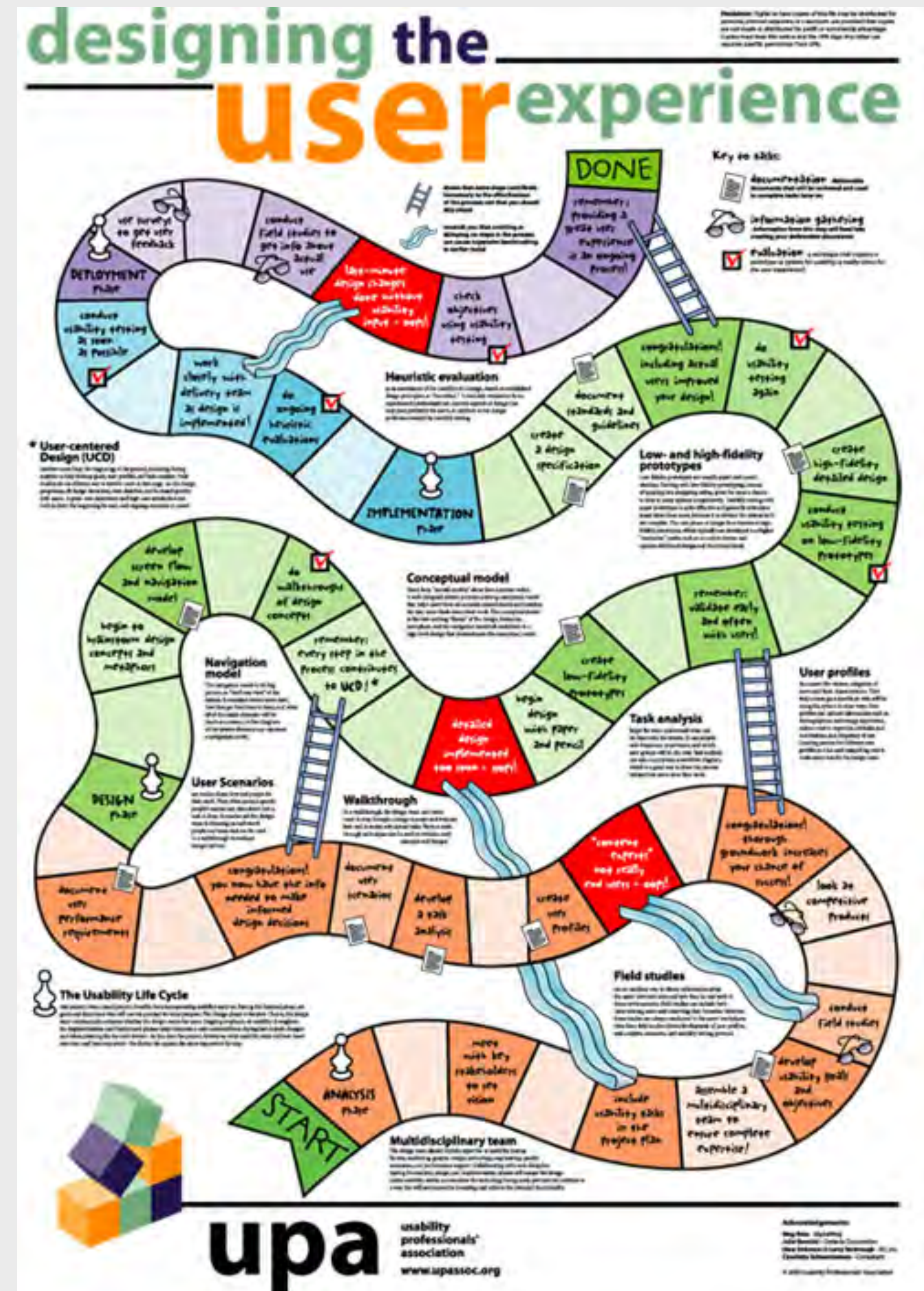


# What is User Centered Design?

The ISO 9241 Standard Describes:

## • Six Principles of Human-Centered Design

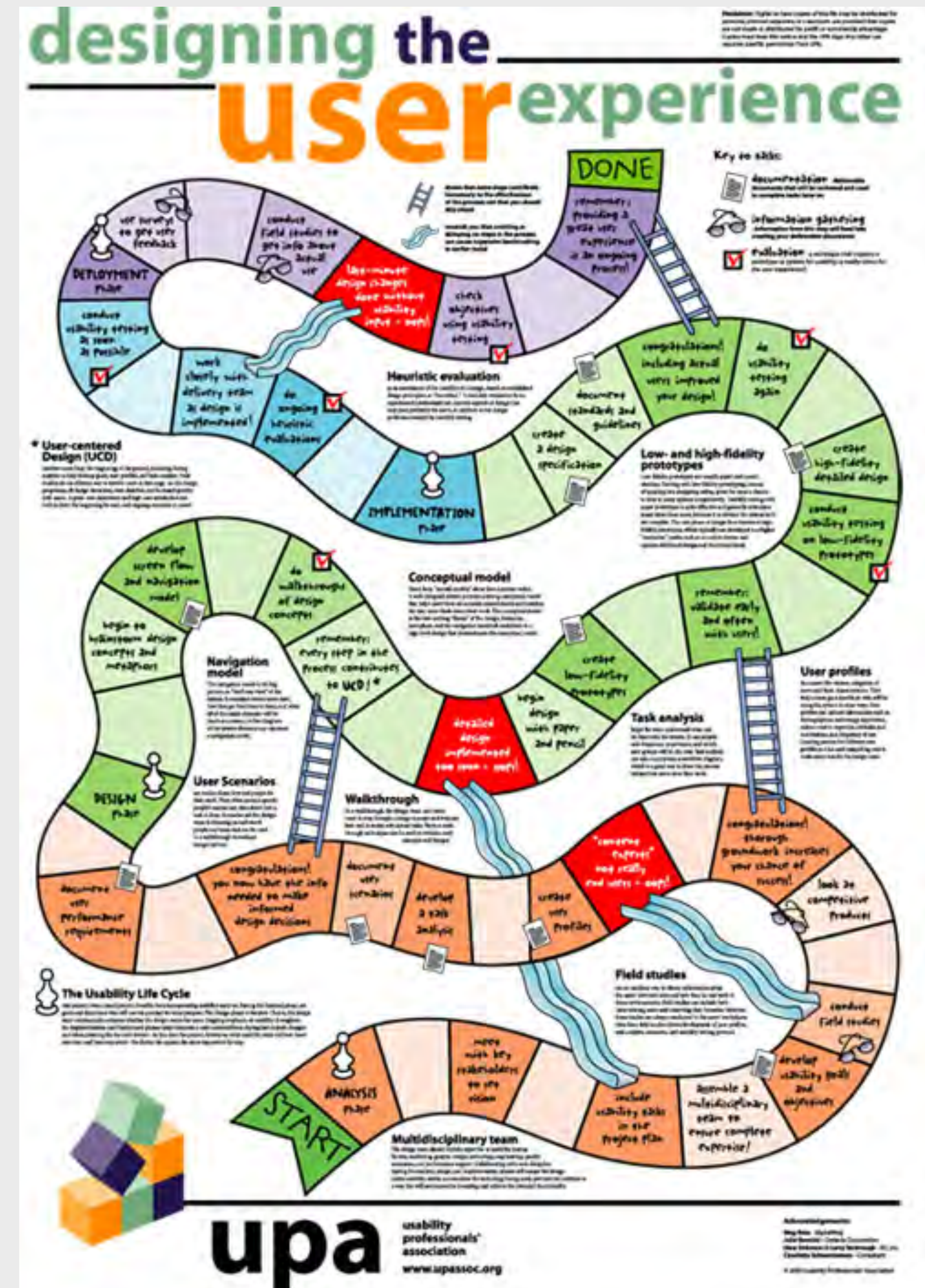
- Design is based upon explicit understanding of user, tasks and environments
- Users are involved throughout design and development
- Design is driven and refined by user-centered evaluation
- Processes are iterative
- Design addresses the whole user experience
- Multidisciplinary design team



# WHAT ACTIVITIES WOULD YOU EXPECT TO FULFILL THESE SIX PRINCIPLES?

## • Six Principles of Human-Centered Design

- Design is based upon explicit understanding of user, tasks and environments
- Users are involved throughout design and development
- Design is driven and refined by user-centered evaluation
- Processes are iterative
- Design addresses the whole user experience
- Multidisciplinary design team



# What is User Centered Design?

... and Four Human-Centered Design Activities:

- **Specify the context of use** (Identify the people who will use the product, what they will use it for, and under what conditions they will use it.)
- **Specify requirements** (Identify any business requirements or user goals that must be met for the product to be successful.)
- **Create design solutions** (This part of the process may be done in stages, building from a rough concept to a complete design.)
- **Evaluate designs** (The most important part of this process is that evaluation - ideally through usability testing with actual users - is as integral as quality testing is to good software development.)

*and...*

The key to developing highly usable products is employing user-centered design.

**TEST**

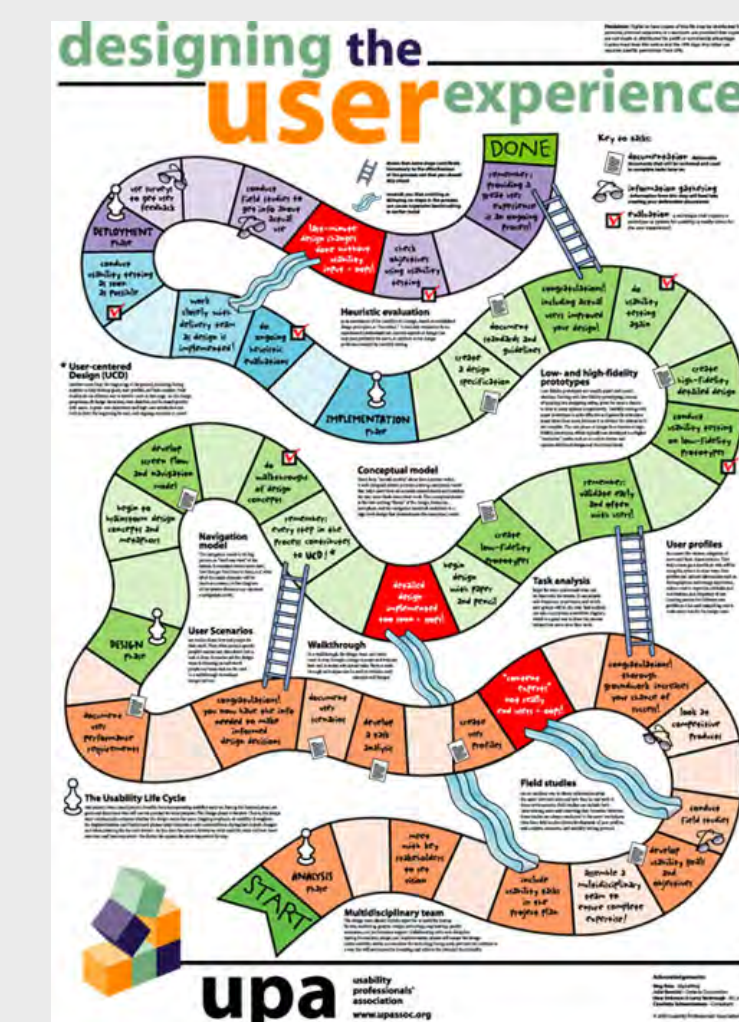
The expression, “test early and often”, is particularly appropriate when it comes to usability testing.

**TEST**

As part of UCD you can and should test as early as possible in the process and the variety of methods available allow you to assist in the development of content, Information architecture, visual design, interaction design and general user satisfaction.

**TEST**

**TEST**





# Typical UCD Methodology

FOUR PHASES (w/suggested activities):

## Analysis Phase

- ▶ Meet with key stakeholders to set vision
- ▶ Include usability tasks in the project plan
- ▶ Assemble a multidisciplinary team to ensure complete expertise
- ▶ Develop usability goals and objectives
- ▶ Conduct field studies
- ▶ Look at competitive products
- ▶ Create user profiles
- ▶ Develop a task analysis
- ▶ Document user scenarios
- ▶ Document user performance requirements

## Design Phase

- ▶ Begin to brainstorm design concepts and metaphors
- ▶ Develop screen flow and navigation model
- ▶ Do walkthroughs of design concepts
- ▶ Begin design with paper and pencil
- ▶ Create low-fidelity prototypes
- ▶ Conduct usability testing on low-fidelity prototypes
- ▶ Create high-fidelity detailed design
- ▶ Do usability testing again
- ▶ Document standards and guidelines
- ▶ Create a design specification

## Implementation Phase

- ▶ Do ongoing heuristic evaluations
- ▶ Work closely with delivery team as design is implemented
- ▶ Conduct usability testing as soon as possible

## Deployment Phase

- ▶ Use surveys to get user feedback
- ▶ Conduct field studies to get info about actual use
- ▶ Check objectives using usability testing

**RESEARCH**  
Measurement

## *What is Usability?*

Usability refers to the quality of a user's experience when interacting with products or systems, including websites, software, devices, or applications.

Usability is about effectiveness, efficiency and the overall satisfaction of the user.

**Usability is not a single, one-dimensional property of a product, system, or user interface.**

## What is Usability?

'Usability' is a combination of factors including:

- Intuitive design: a nearly effortless understanding of the architecture and navigation of the site
- Ease of learning: How fast a user who has never seen the user interface before can accomplish basic tasks
- Efficiency of use: How fast an experienced user can accomplish tasks
- Memorability: after visiting the site, if a user can remember enough to use it effectively in future visits
- Error frequency and severity: how often users make errors while using the system, how serious the errors are, and how users recover from the errors
- Subjective satisfaction: If the user likes using the system

Measurement

# ***Why Would You Conduct a Usability Test?***

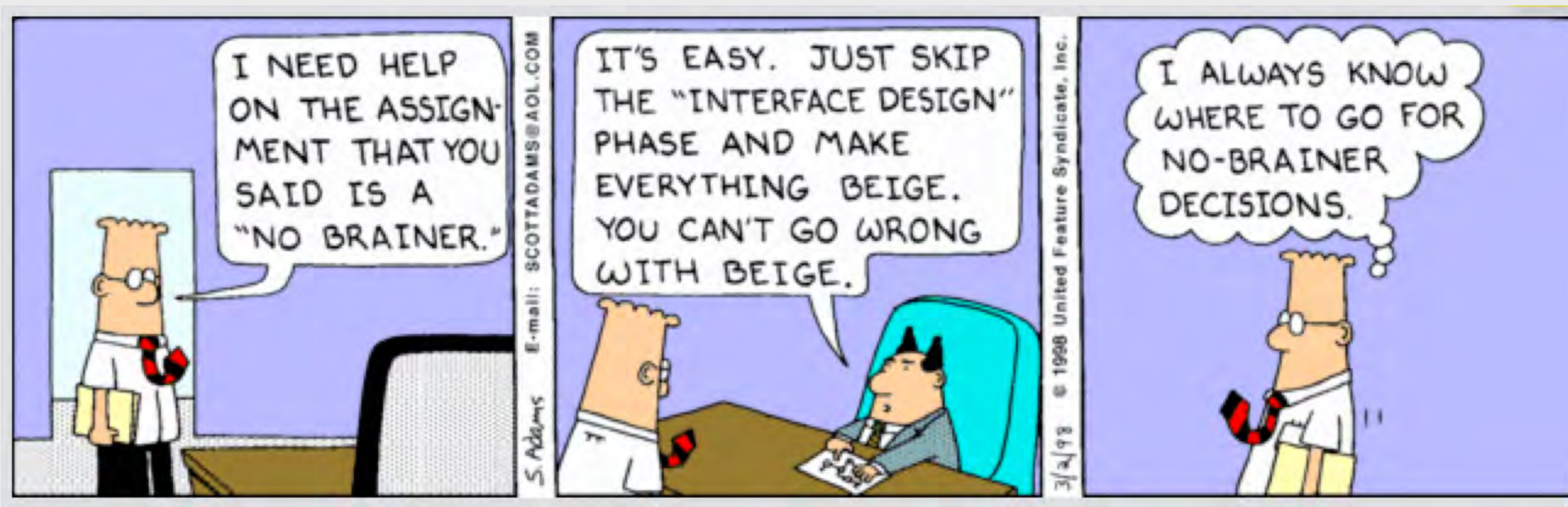
DISCUSS.....

# *Why Are You Conducting a Usability Test?*

- Checking to see if the design works
- Diagnosing problems
- Comparing Alternative
- Verifying that design goals are met

## Benefits of Usability Testing

- Gets feedback directly from users - not intermediaries



# Benefits of Usability Testing

- Provides Data (not opinions) for design decisions

“Make it look like ‘American Express’” - Citibank executive in response to a UI design

“Go back and change it in the next release” - Citibank executive in response to his previous decision

“It’s going to take two releases to change it back” - Project Management to Citibank executive

“I think the users would like the color red... - ex Verizon Manager”

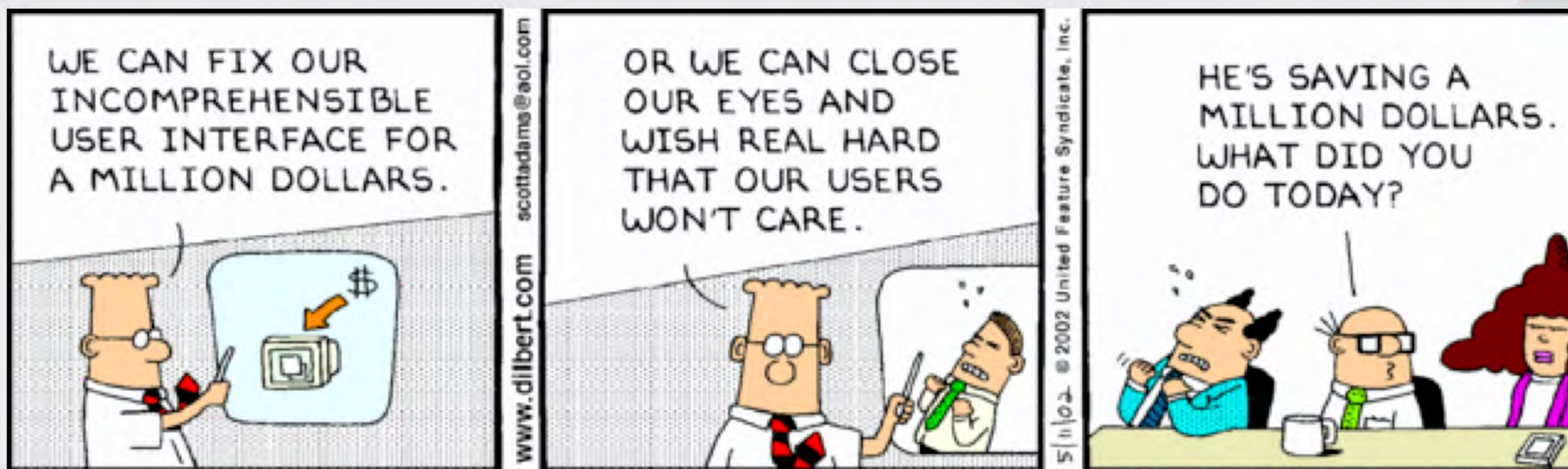
“I don’t like the Terms & Conditions at the bottom of the Page, put it in the middle” - Project Director



**ACTUAL QUOTES FROM  
REAL PEOPLE!**

## Benefits of Usability Testing

- Saves development time by avoiding extensive rework late in the development process
- Creates a positive return on investment (ROI)

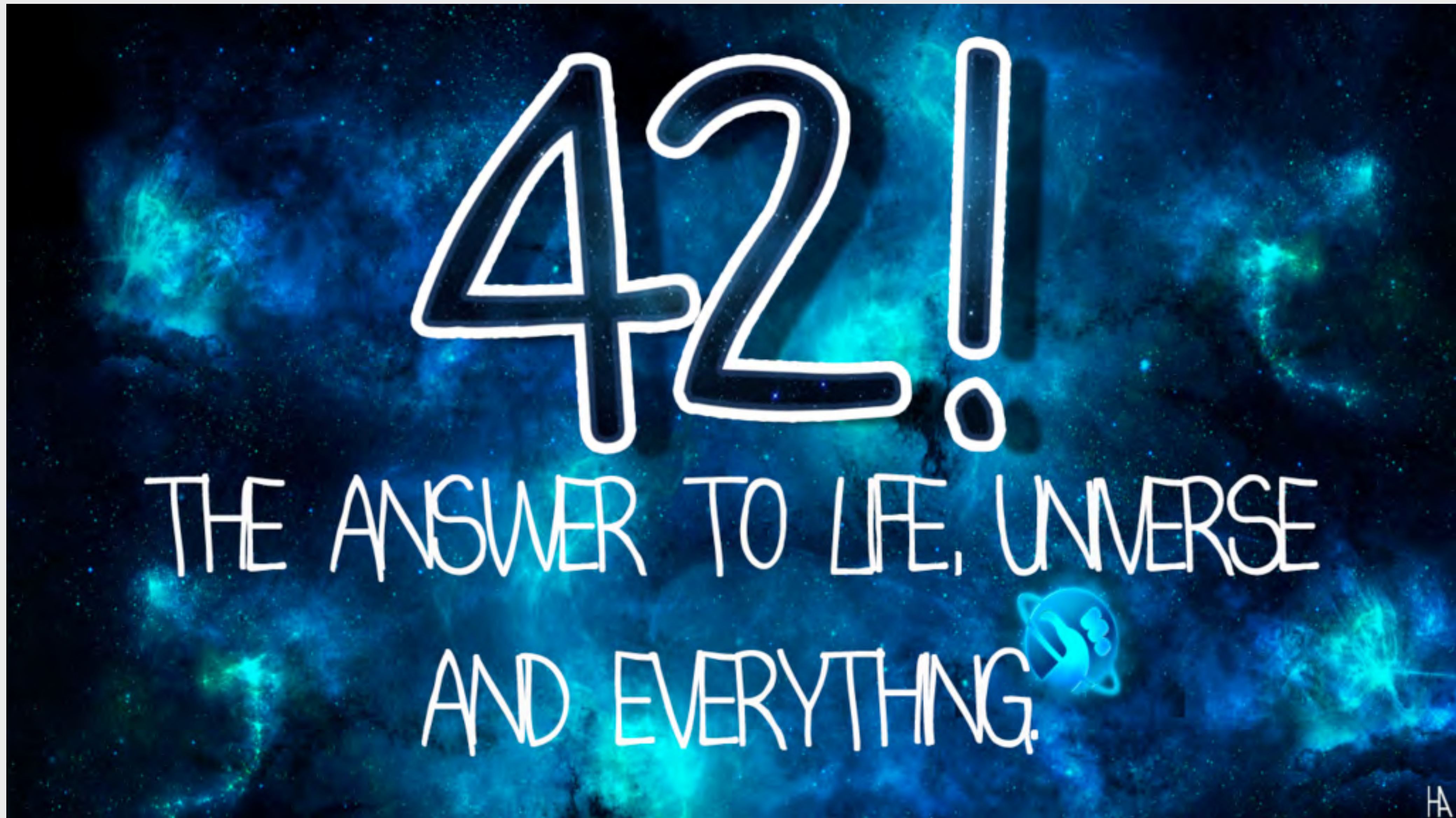




# Why Performance Tests are Critical

- Subjective ratings can be misleading
  - People often report satisfaction with products that are poorly designed
  - Problems are typically downplayed by users
- Subjective ratings can be influenced
  - “Hot” factor
  - “Sympathy” factor **How might these effect subjective ratings?**
  - “Willingness to Please”
  - Perceived completion level of what’s being tested
    - Low Fidelity vs High Fidelity **How might this effect subjective ratings?**

## How do we Measure all of this stuff?



Unfortunately, no....this won't work.....

# How do we Measure all of this stuff?

## Qualitative vs Quantitative Data

- **Quantitative**

- Anything that can be expressed as a number, or quantified
- These data may be represented by ordinal, interval or ratio scales

- **Qualitative**

- Data that cannot naturally be expressed as a number.
- Examples include gender, socio economic status, feelings, perceptions, etc.

OK to compute....	Nominal	Ordinal	Interval	Ratio
frequency distribution.	Yes	Yes	Yes	Yes
median and percentiles.	No	Yes	Yes	Yes
add or subtract.	No	No	Yes	Yes
mean, standard deviation, standard error of the mean.	No	No	Yes	Yes
ratio, or coefficient of variation.	No	No	No	Yes

# *How do we Measure all of this stuff?*

## *With Data Types...*

- **Nominal Data**

- A categorical variable, also called a nominal variable, is for mutual exclusive, but not ordered, categories. (Blonds, Brunets, Red Heads; Chevy, Ford, Nissan)
- It is generally inappropriate to compute the mean for Nominal variables

**Suppose you had 20 subjects, 12 of which were male, and 8 of which were female.**

**If you assigned males a value of '1' and females a value of '2', could you compute the mean sex of subjects in your sample?**

**It is possible to compute a mean value, but how meaningful would that be?**

**How would you interpret a mean sex of 1.4?**

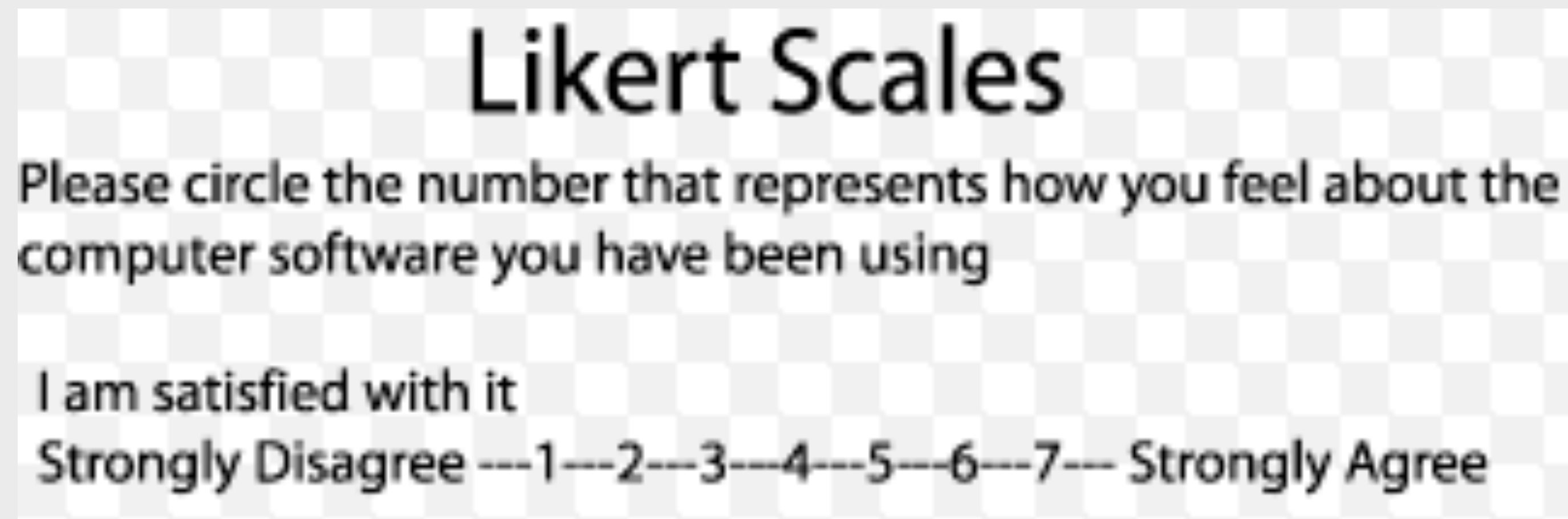
**When you are examining a Nominal variable such as sex, it is more appropriate to compute a statistic such as a percentage (60% of the sample was male).**

# How do we Measure all of this stuff?

## With Data Types...

- **Ordinal Data**

- A ordinal variable, is one where the order matters but not the difference between values.
- Psychometric Research shows us that the human psychological distance between 6 and 7 on a Likert scale is much greater than between a 3 and 4.



# How do we Measure all of this stuff?

## More Data Types...

- **Interval**

- A interval variable is a measurement where the difference between two values is meaningful.
- The difference between a temperature of 100 degrees (F) and 90 degrees is the same difference as between 90 degrees and 80 degrees.

- **Ratio**

- A ratio variable, has all the properties of an interval variable, and also has a clear definition of 0.0. When the variable equals 0.0, there is none of that variable.
- Variables like height, weight, 0 degrees Kelvin are ratio variables.

OK to compute....	Nominal	Ordinal	Interval	Ratio
frequency distribution.	Yes	Yes	Yes	Yes
median and percentiles.	No	Yes	Yes	Yes
add or subtract.	No	No	Yes	Yes
mean, standard deviation, standard error of the mean.	No	No	Yes	Yes
ratio, or coefficient of variation.	No	No	No	Yes

# *How do we Measure all of this stuff? With Data Types...*

## **Ordinal Data Issue**

In the social and behavioral sciences, much of what we study is measured on what would be classified as an ordinal level. We often ask if people "Strongly Disagree", "Slightly Disagree", or are "Neutral" to a series of statements. We then assign a value of '1' if they Strongly Disagree with a statement, up to a '5' if they Strongly Agree with a statement. This type of measurement is ordinal, in the sense that "Strongly Agree" reflects more agreement than "Slightly Agree". This type of measurement is not an interval or a ratio level of measurement, because we can not state for certain that the interval between "Strongly Disagree" and "Slightly Disagree" is equivalent to the interval between "Slightly Disagree" and "Neutral". Nor can we say that there is an absolute zero point for level of agreement. However, if we were to rigidly follow the rules of "permissible" analyses for ordinal variables, many of the analyses we conduct in social sciences research would be deemed impermissible. On the other hand, some scientists have conducted computer simulations to try and find out what would happen if we violated certain "rules" of data analysis in specific cases. They have found that **SOMETIMES**, it is alright to treat ordinal data (such as variables which have been measured using Strongly Disagree to Strongly Agree response alternatives) as though it were interval level data, and conduct statistical tests that are appropriate for interval level data.

# How do we Measure all of this stuff?

## Descriptive vs Inferential Statistics

Before discussing the differences between descriptive and inferential statistics, we must first be familiar with two important concepts in social science statistics: population and sample.

- A population is the **total set** of individuals, groups, objects, or events that the researcher is studying.
- A sample is a relatively **small subset** of people, objects, groups, or events, that is selected from the population.

### • Descriptive Statistics

- Data could be collected from either a sample or a population
- Can **ONLY** be used to **DESCRIBE** the group that is being studied
- Frequency distributions, Mean (average), Median (the value that divides the distribution in half), Mode (the value that occurs the highest number of times) and graphs like pie/bar charts are examples of descriptive statistics
- “9 out of 10 Dentists surveyed prefer Trident sugarless gum”
  - (but we only surveyed 10 Dentists)

### • Inferential Statistics

- Data is generally collected from a sample of the population being studied
- The sample must be representative enough that any results can be generalized from the sample to the whole population
- Tests of significance are used to assess the probability that the results of the analysis could have occurred by chance
- You must have a sufficiently large sample size that provides you with statistical significance
- Tests of significance entail the use of hypotheses



## So Keep In Mind.....

- **Statistically Significant Results** are a result of the FORMAL SCIENTIFIC METHOD with all the goodies of  $H_0$  (null hypothesis) and  $H_a$  (alternate hypothesis) , correct population sampling & screening and formal testing methods that eliminate bias.
  - Even with statistically significant results you can only say that “the data do indicate that...” and/or “There was a failure to reject the Null Hypothesis / The Null Hypothesis was rejected in favor of...”
  - You NEVER “PROVE” anything because “PROOF” comes from testing the ENTIRE POPULATION and NOT a Sample.
- **Descriptive Statistics** (Mean, Median, Mode) provide some trend data and is valuable to point you in the direction of inquiry, but remember “9 out of 10 Dentists surveyed prefer that their patients chew Trident gum.”
- **Correlation does NOT infer Causation** - Just because you see one thing increase while another decreases does not mean that one causes the other. Further exploration is needed.

## *So what will you use.....*

### *IT DEPENDS....*

- If you do any testing using software packages like 'Optimal Workshop', 'UserZoom', etc they will do the analysis for you
  - You MUST make sure that the Sample being tested (people) are Representative of the TOTAL population
  - This covers card sorts, tree tests, first click tests
  - No fair using anyone associated with the project
- If you do Heuristic Analyses (UserFocus) the spreadsheet scores for you
  - You must use 5 or more (Usability Professionals) for representative results
  - You and your team can use the UserFocus Methodology as a checklist for a good design
    - We will discuss in depth later in this course
- ASK US .....

# Summary

- True Research uses the Scientific Method
- User Centered Design (UCD) is a PROCESS based on an International Standard in which users and data drive the design
- User Research is an integral part of the Iterative process of good design
- User Research allows you to
  - Check to see if the design works
  - Diagnose problems
  - Compare Alternatives
  - Verify that design goals are met
- Knowing your Data is critical
  - The specific type of data you have will dictate the types of statistics you can use (or can't use) and how you REPORT them
- Qual vs Quan
  - Qualitative Data can be observed but not measured (smell, color, texture, taste, feelings)\*\*
  - Quantitative Data deals with number and can be measured (length, height, area, time, speed, cost, age)
- Data Types
  - Nominal - You can name it but that's about all (Blondes, Brunets, Redheads)
  - Ordinal - The order matters but not the interval between the placements (1st born, 2nd born, 3rd born)
  - Interval - The measurement between the two values is meaningful (there is 75 cents difference between a quarter and one dollar)
  - Ratio - Interval but with an absolute zero (\$100 is exactly 100 times \$1)
- Statistical Types
  - Descriptive - Summarizes a SAMPLE (a subset of the population)
  - Inferential - Makes predictions about a larger population that the sample represents; used to test hypotheses and make estimations. (w/a sample size that allows statistical inference)

*and.....*

**DON'T PANIC**



# Basics

## Graphical Representation

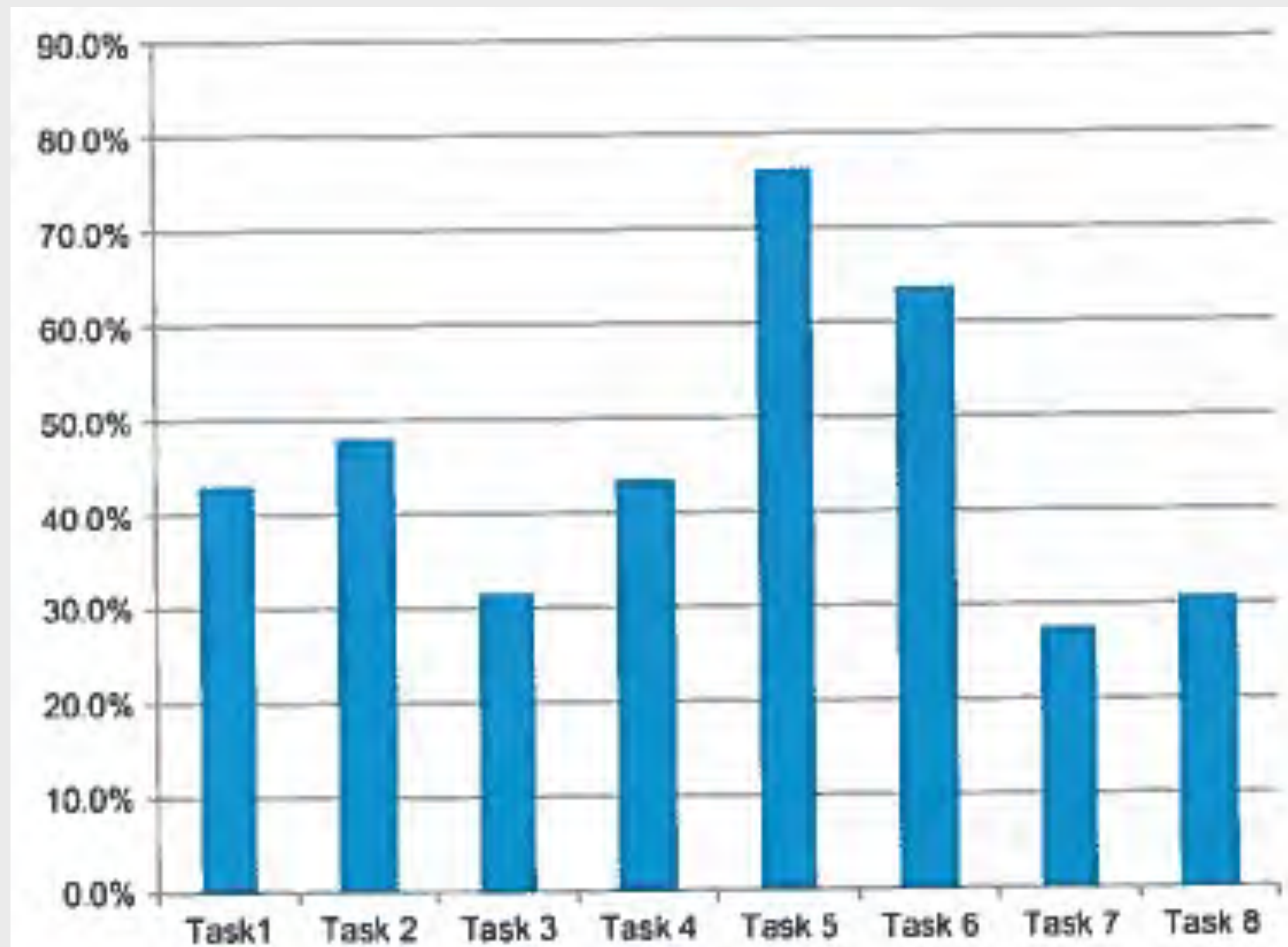
August 2014

# Data Graphs - Data is no good if you can't communicate the results to people

- Five Basic Types of Data Graphs
  - Column or bar graphs
  - Line graphs
  - Scatterplots
  - Pie or donut charts
  - stacked bar or column graphs
- General Considerations
  - Completely label the axes and units
    - *Is it minutes, hours or days??*
  - Don't imply more precision in your data than it deserves
    - *Did you really measure time down to the hundredths of a second?*
  - Don't use color ALONE to convey information
    - *Some people are color blind*
  - If available, show confidence intervals
    - *Use this for means of participant data (times, ratings, etc)*
    - *This show variability (fluctuations) in the data*
  - Don't overload your graphs
    - *Just because you CAN put everything in a single graph doesn't mean you should*
  - Be careful with 3D graphs
    - *Ask yourself does it really help understand the data or is it just 'cool'.*

## Column or Bar Graphs

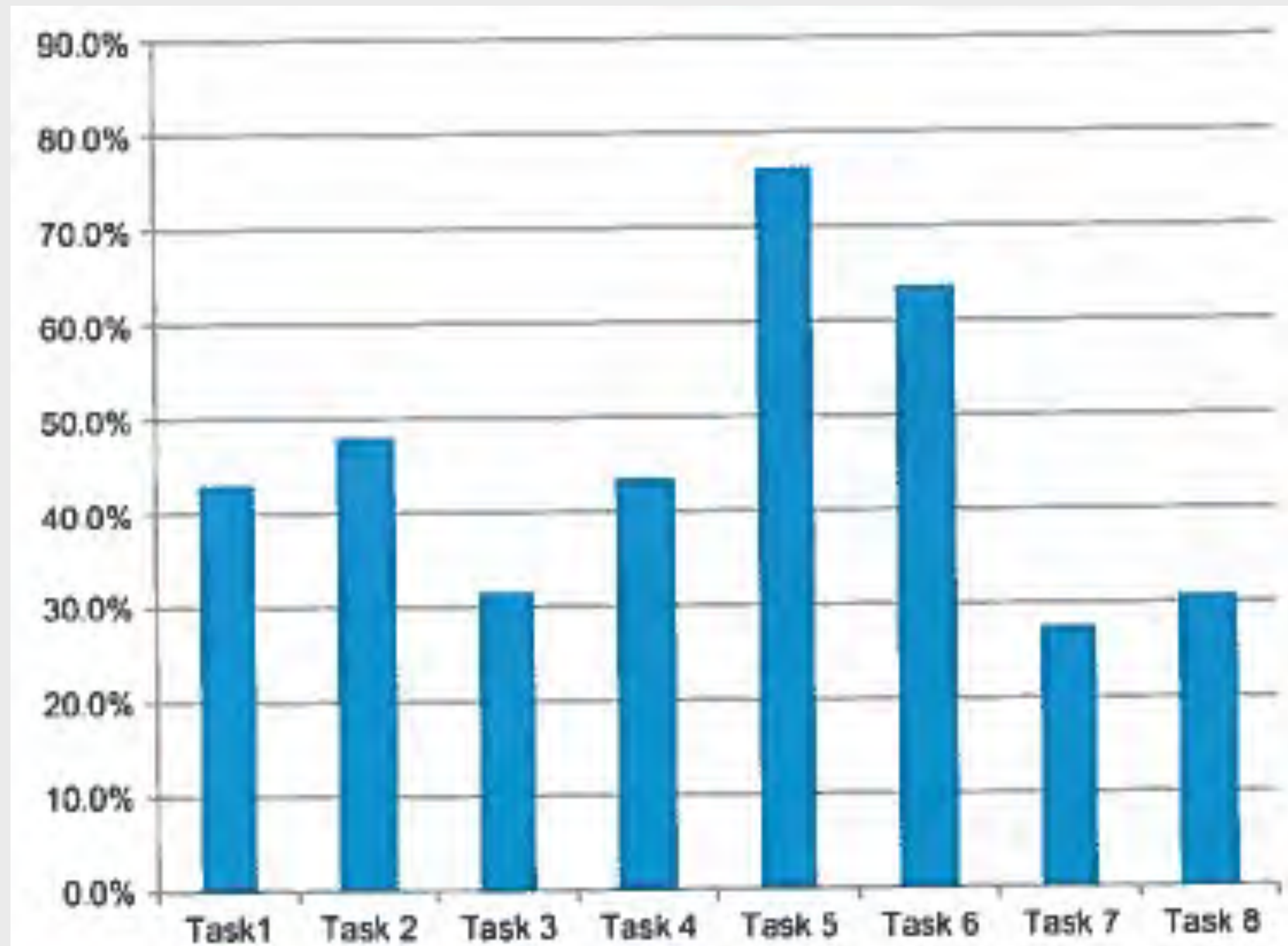
- Column is Vertical; Bar is Horizontal
- One of the most common ways to present usability data
- Appropriate when you want to present the values of times, percentages, etc (continuous data) for discrete items or categories (tasks, participants, designs)



**What is WRONG with the chart on the left?**

## Column or Bar Graphs

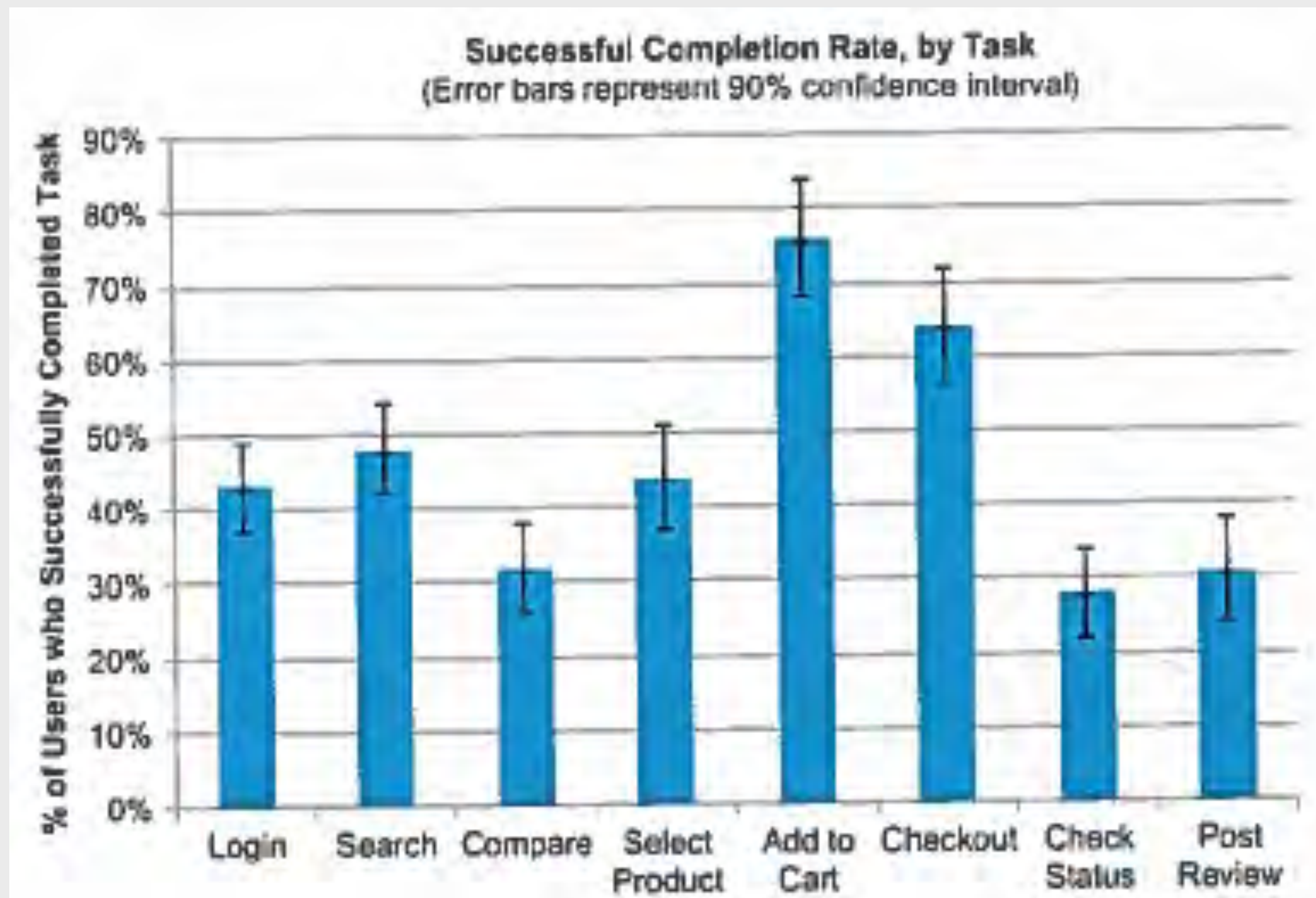
- Failing to label data
- Not showing confidence intervals when you can
- Showing too much precision in vertical axis labels





## Column or Bar Graphs

- Should Look like this



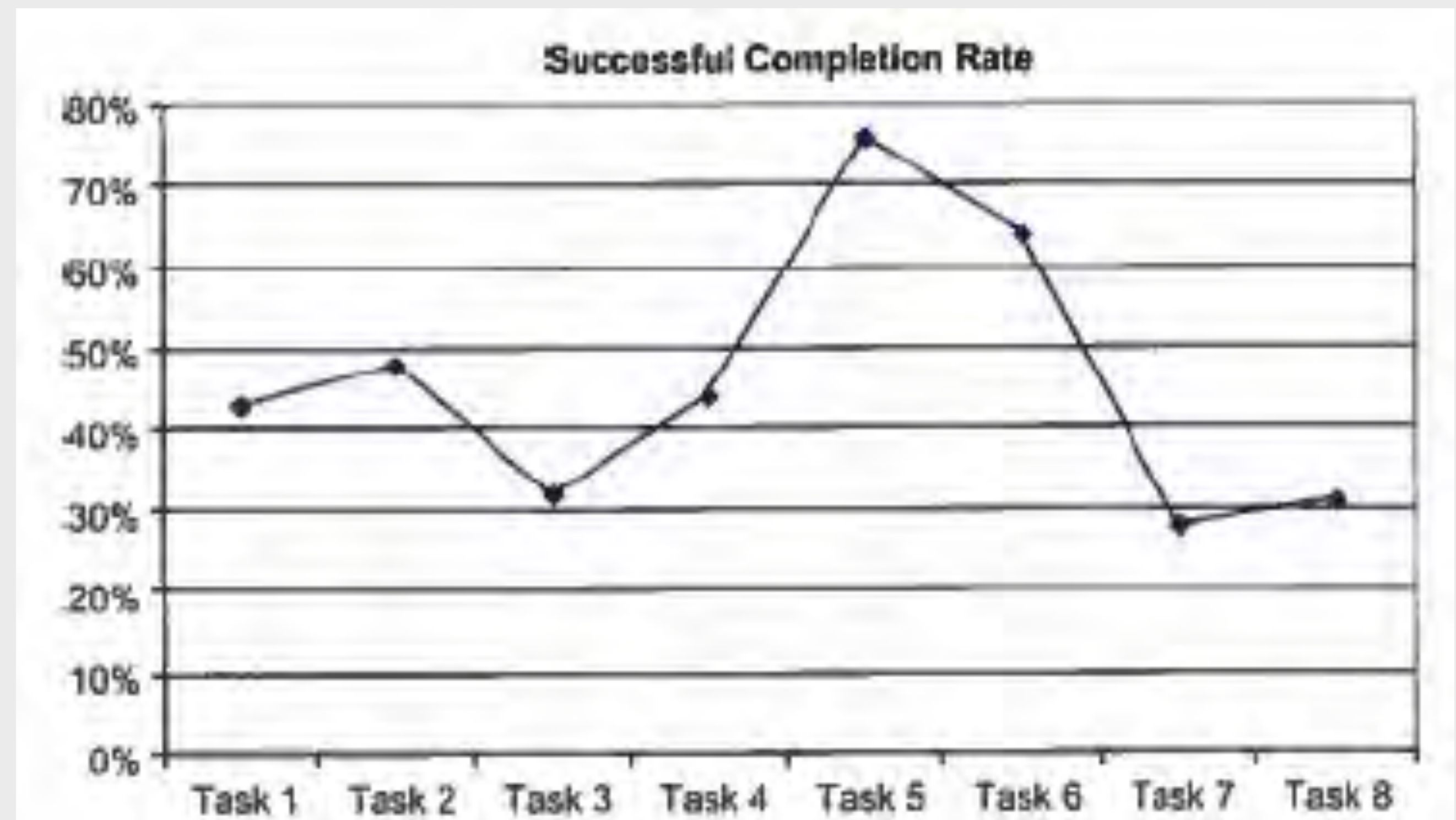
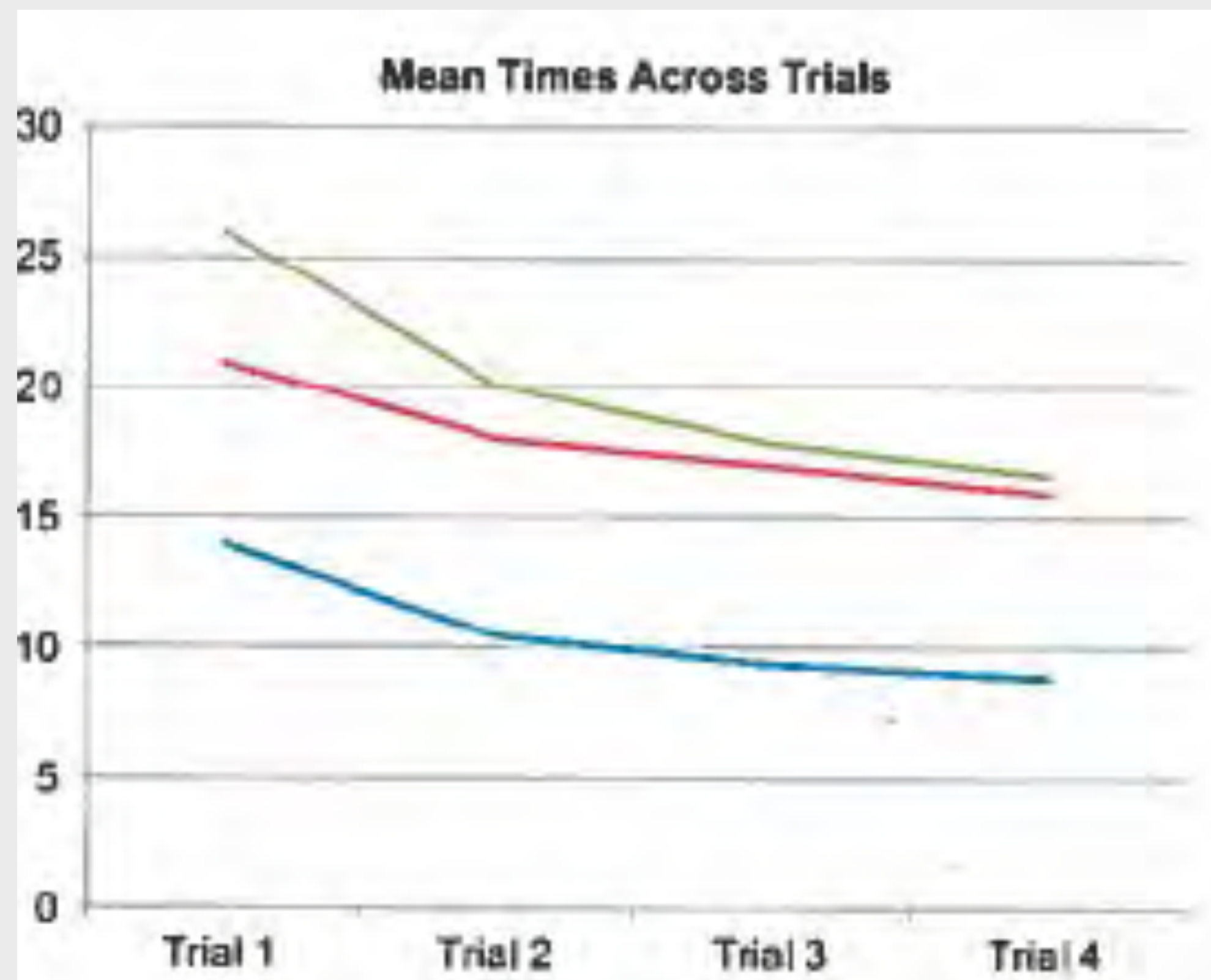
Confidence Interval -

- a range of values so defined that there is a specified probability that the value of a parameter lies within it.
- a measure of the reliability of an estimate

# Line Graphs

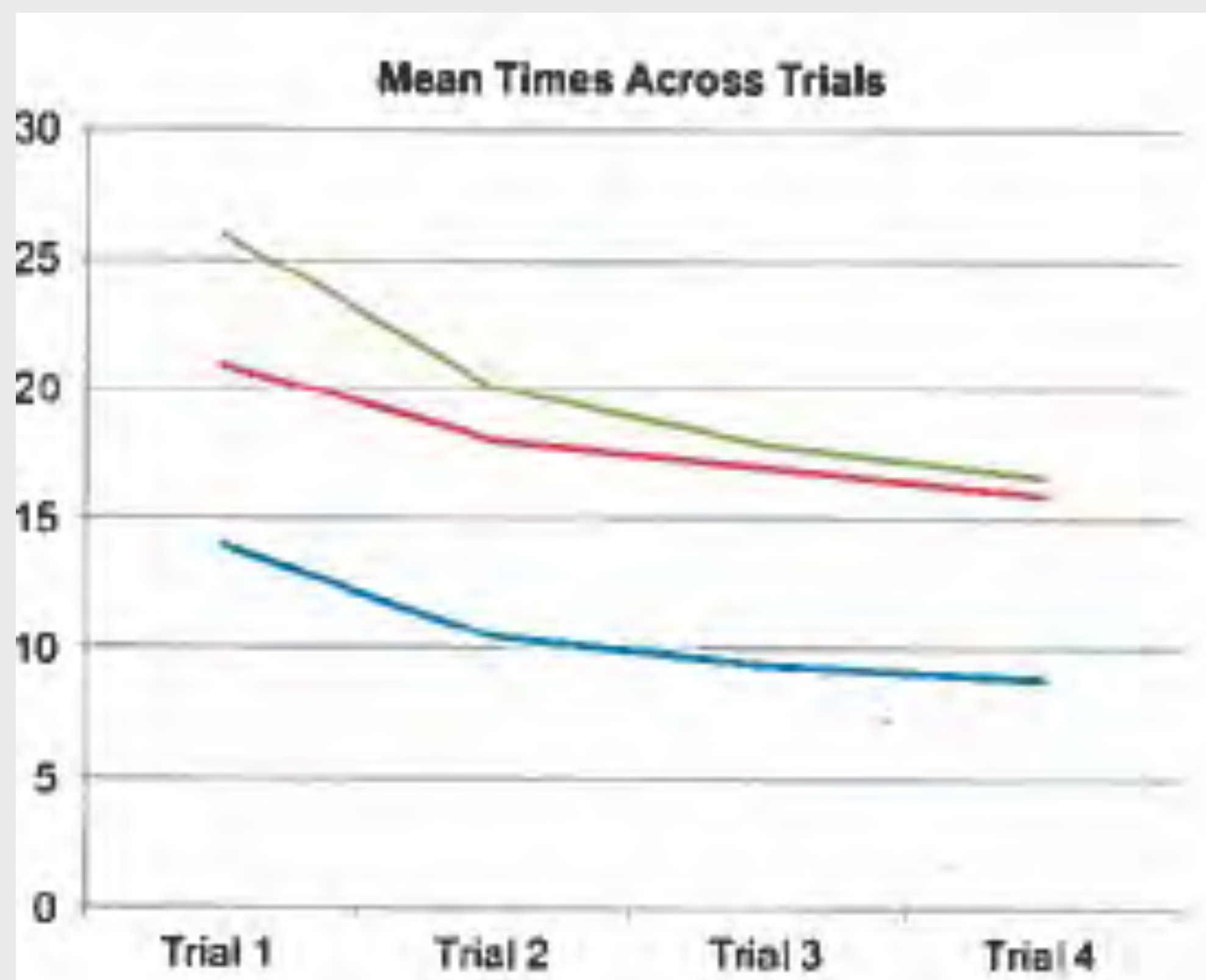
- Appropriate when you want to show values of one continuous variable (% correct, # of errors) as a function of another continuous variable (age, time, trial of a test)
- IF one of the variables is discrete (gender, name, model of car) then use a bar graph
- Show your data points connected by lines to point out trends
- Include a Legend if you have more than one line

**Exercise: What is WRONG with these charts?**

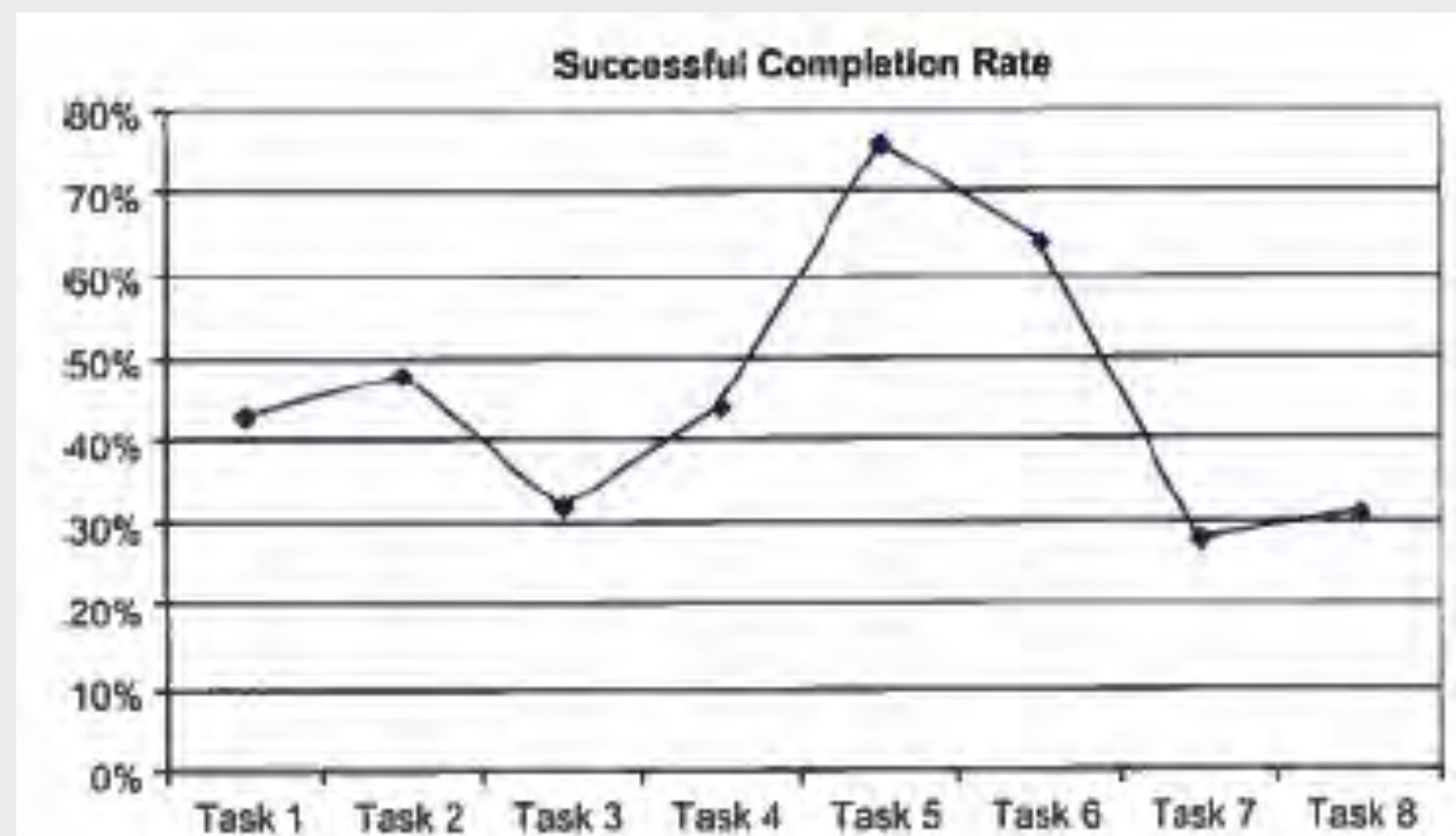


## Line Graphs

- Failing to label the vertical axis
- Not showing data points
- Not including a legend
- Not showing Confidence Intervals

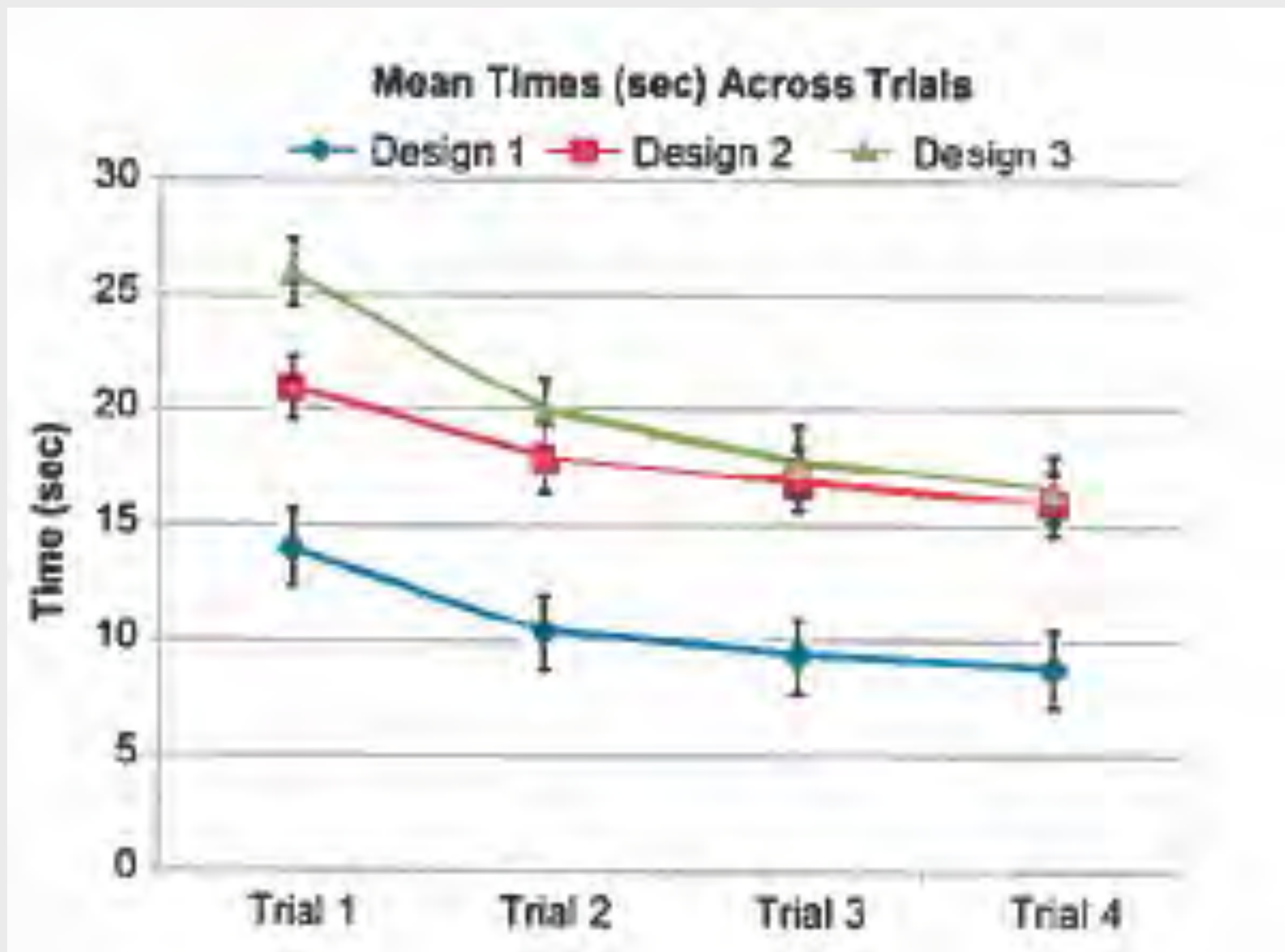


- Lines imply that the tasks are a continuous variable and they are NOT



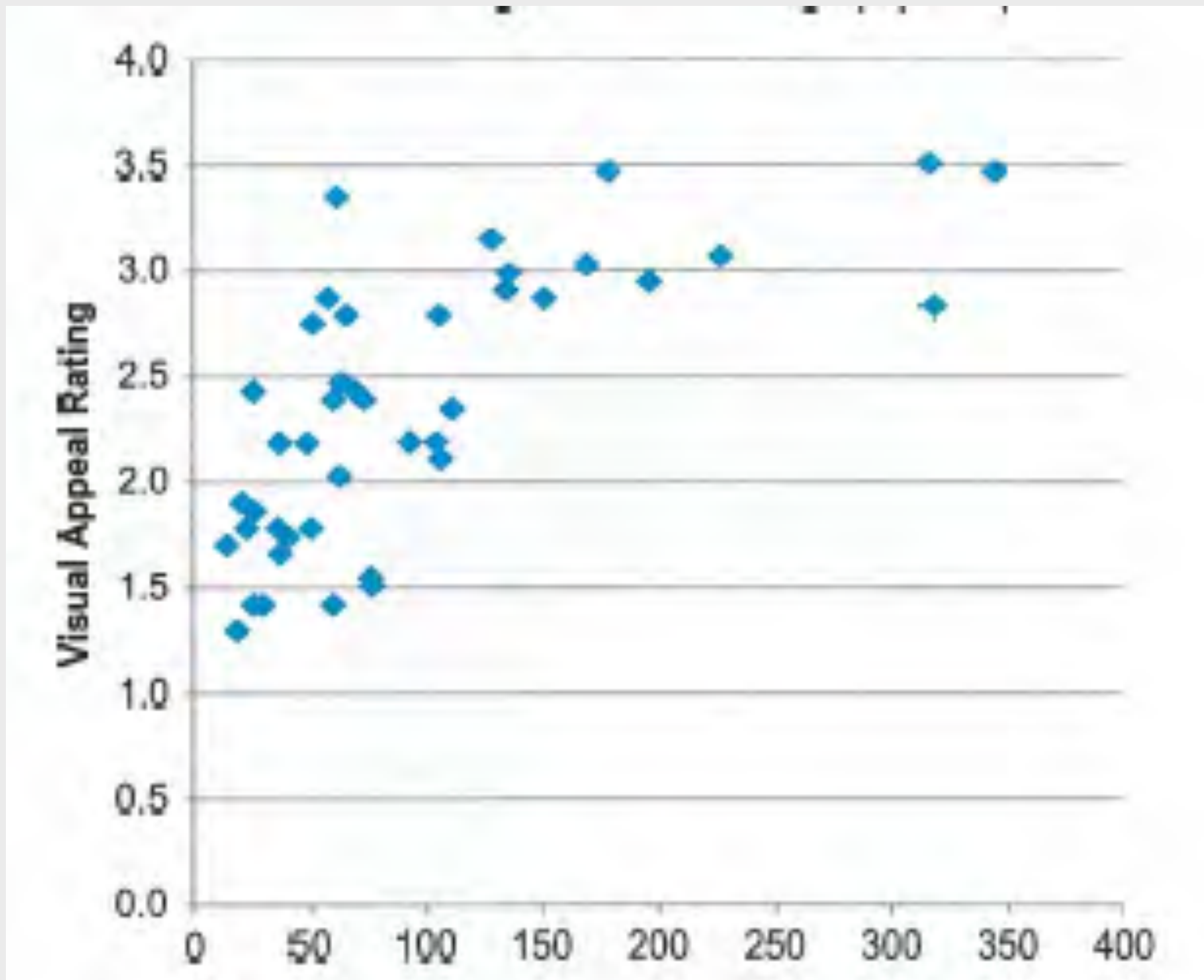
## Line Graphs

Should look like this



# Scatterplots

- When you have paired values you want to plot
- Usually both variables would be continuous
- Used to illustrate a relationship between two variables
- Often helpful to add a trend line

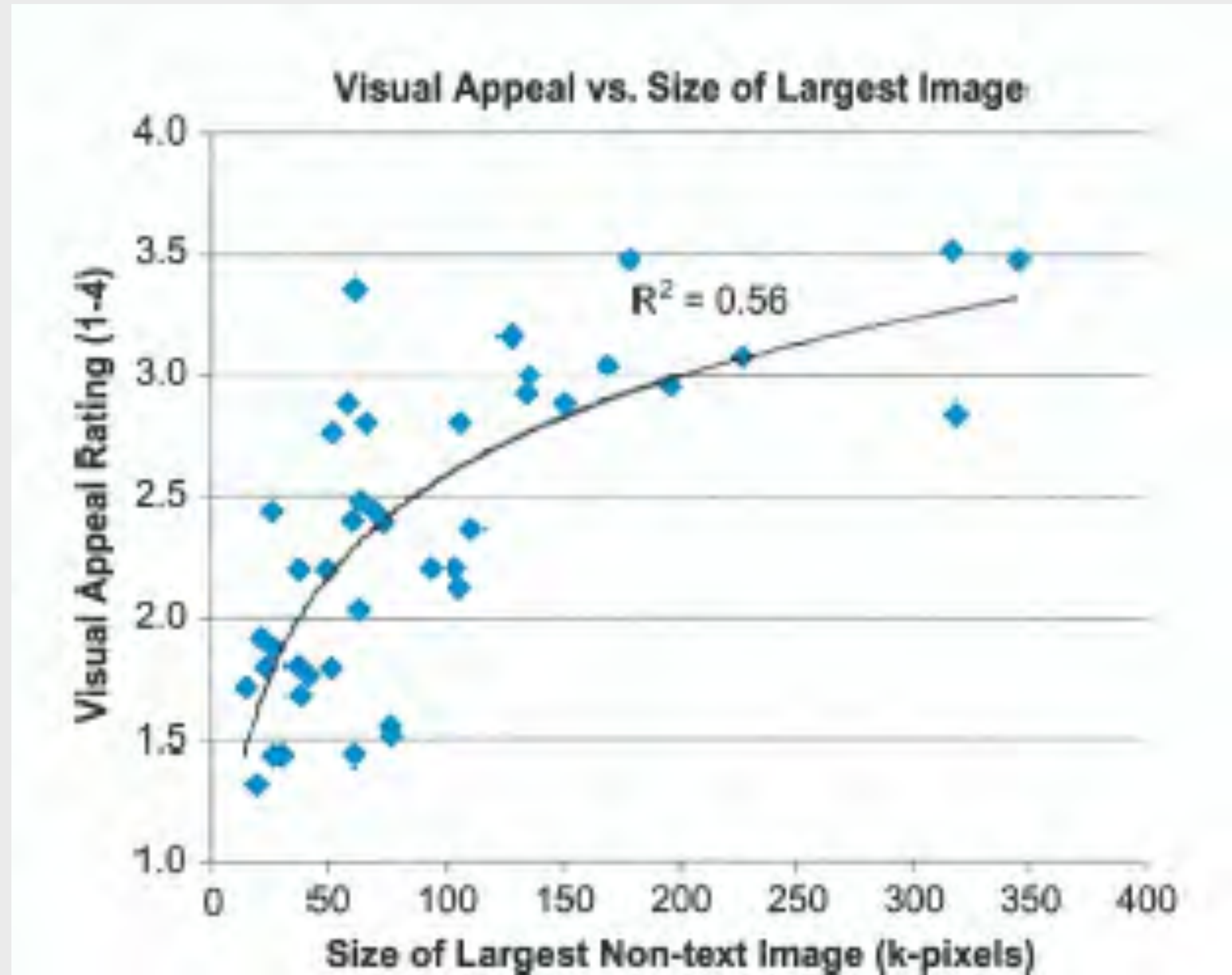


**Exercise:**

**What is WRONG with the chart on the left?**

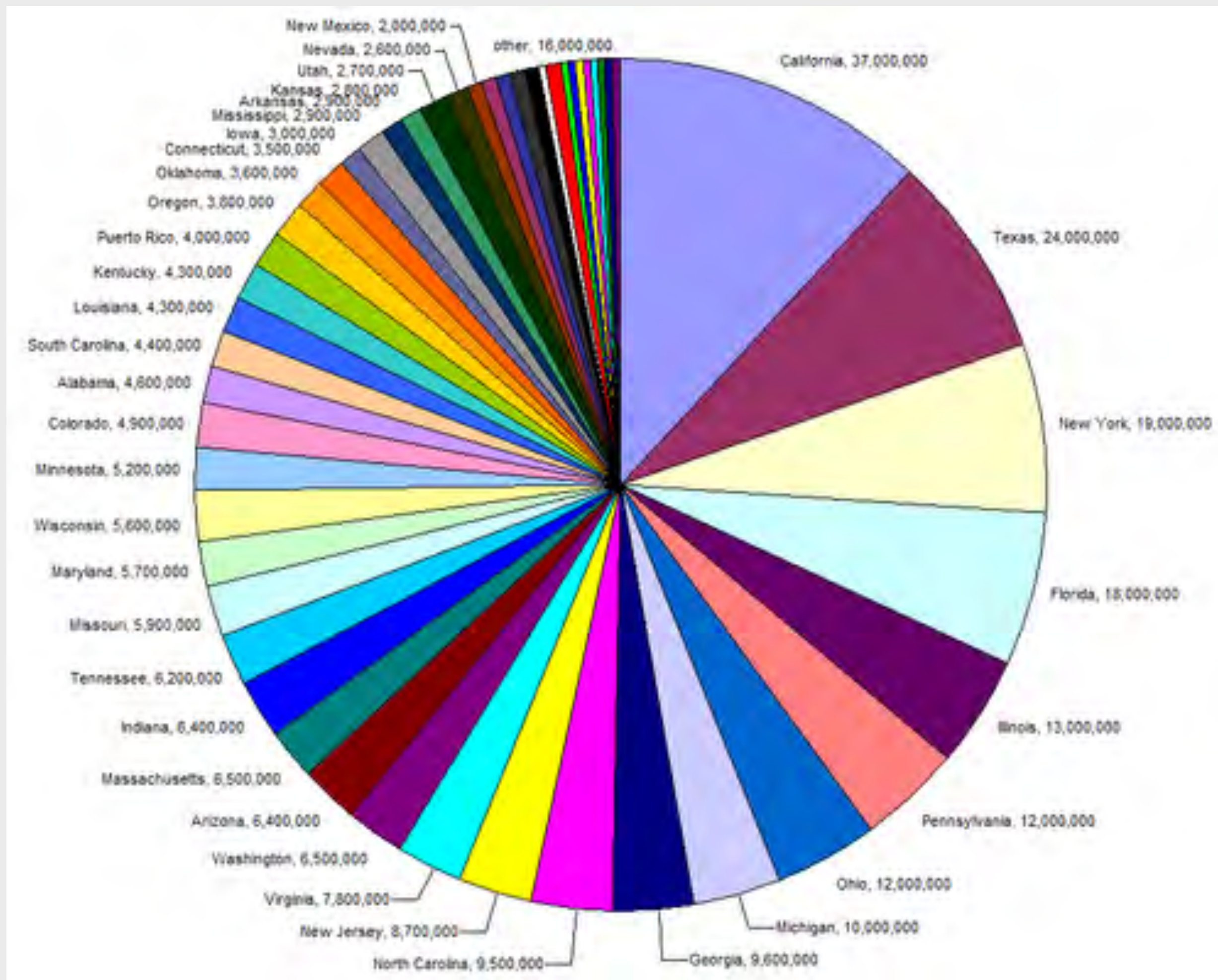
# Scatterplots

What it should look like



## Pie and Donut Charts

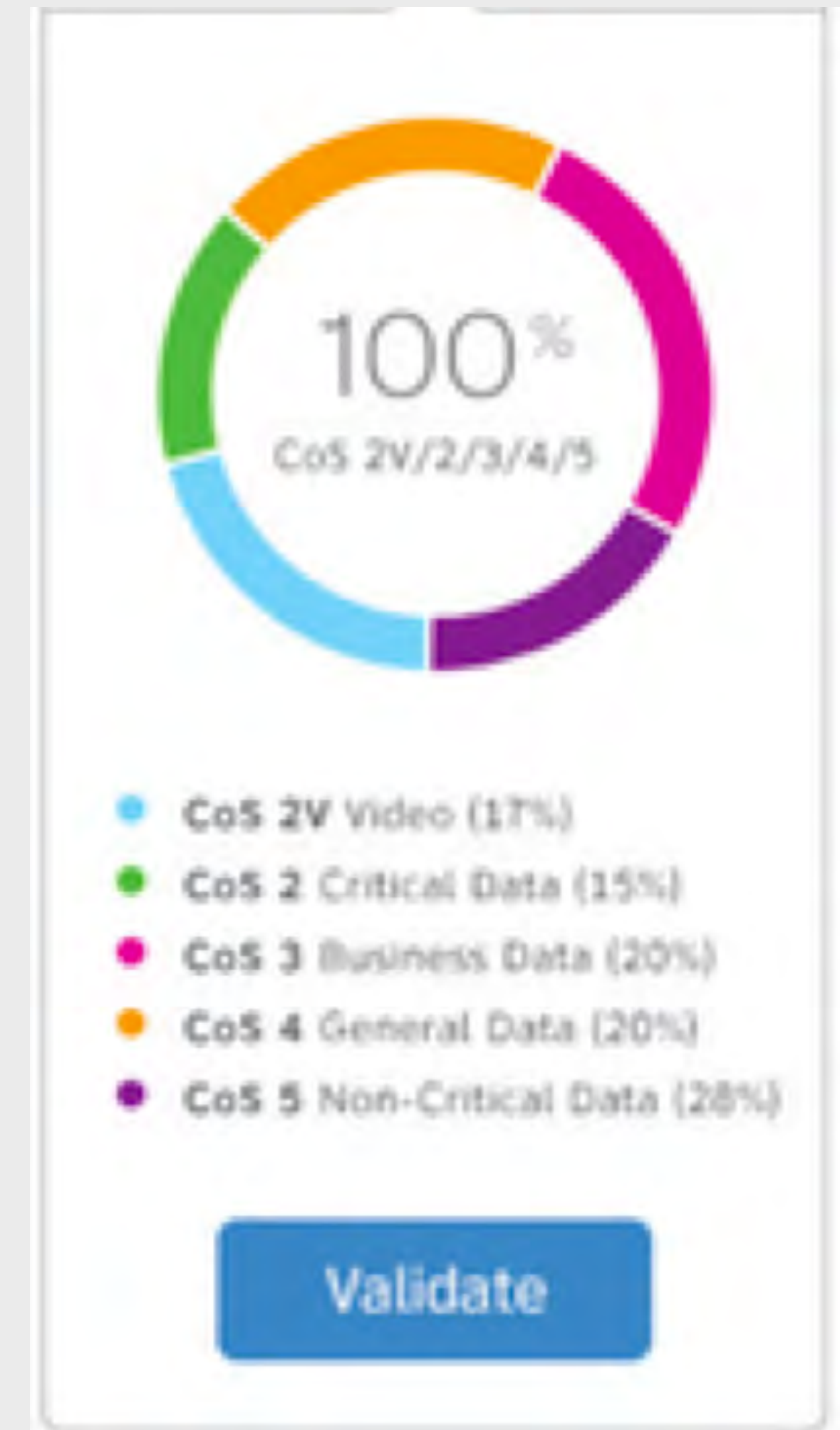
- Illustrate the parts or percentages of a **WHOLE** (pies and donuts are whole until you divide them up)
- Never have more segments than can easily convey the information to the user (keep to < 7)
- Include % and label for each segment
- Order the data meaningfully



### Exercise:

What is **WRONG** with the chart on the left?

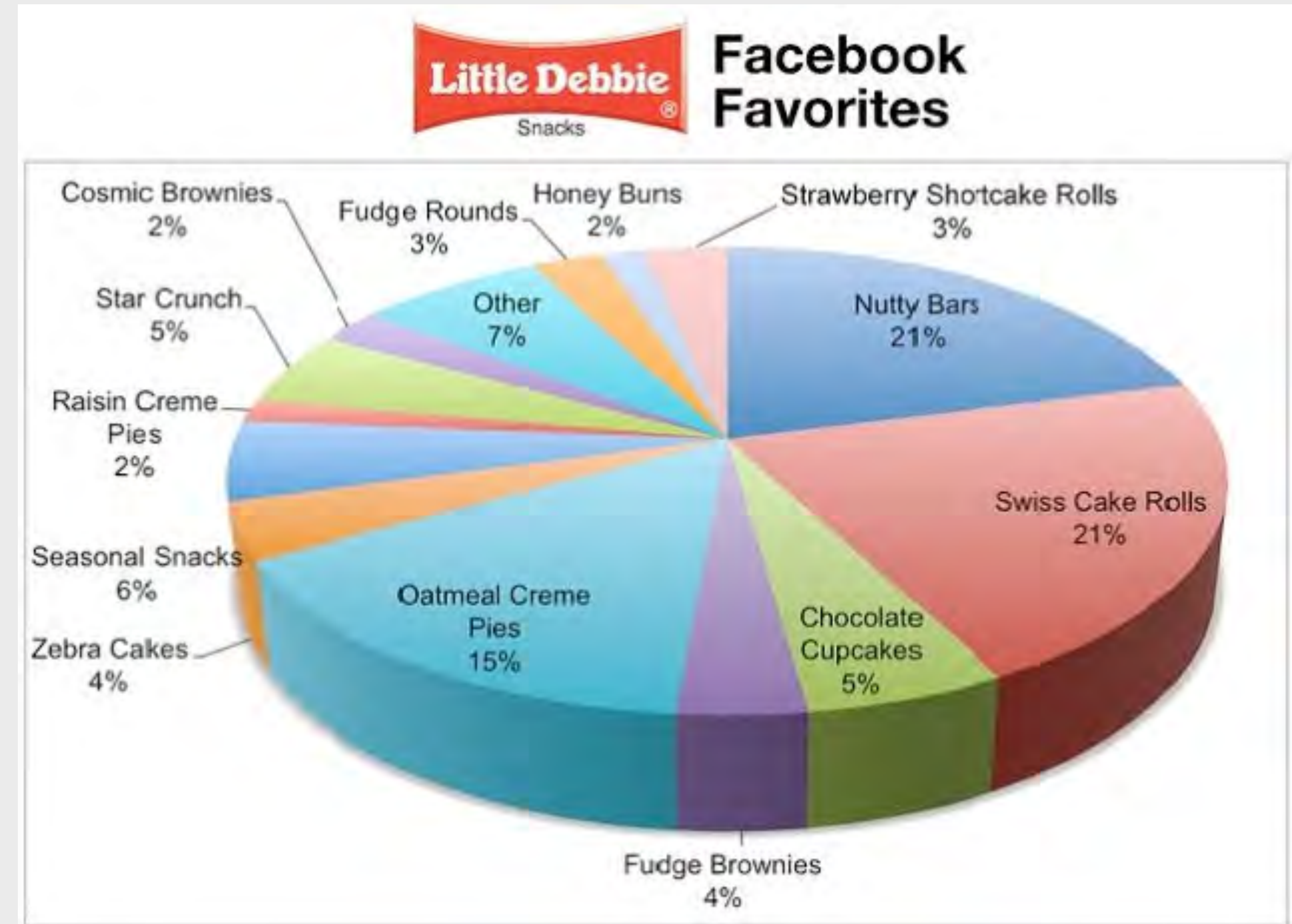
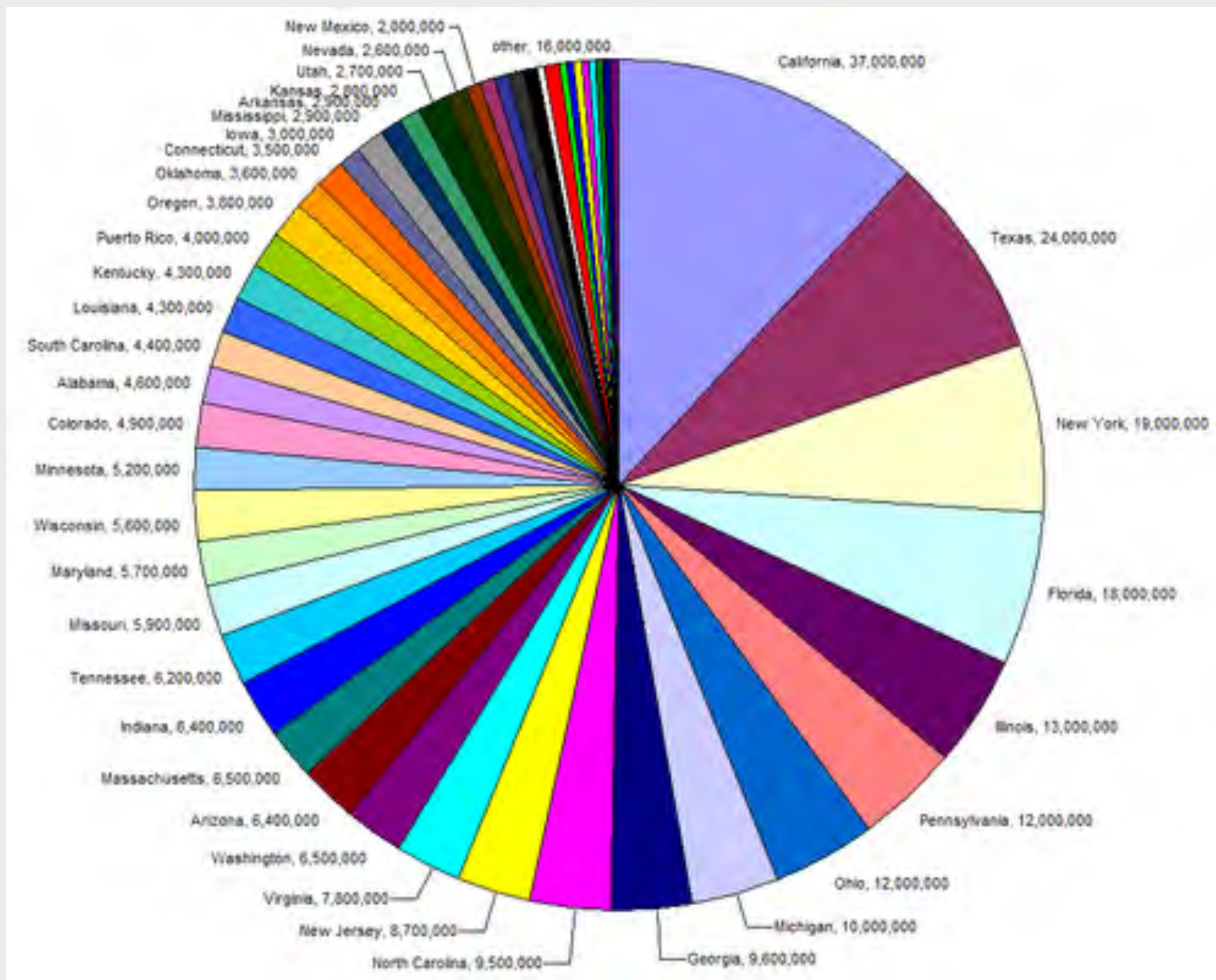
On the Right?



## Pie and Donut Charts

- Too many segments
- Not expressed as a percentage
- No legend

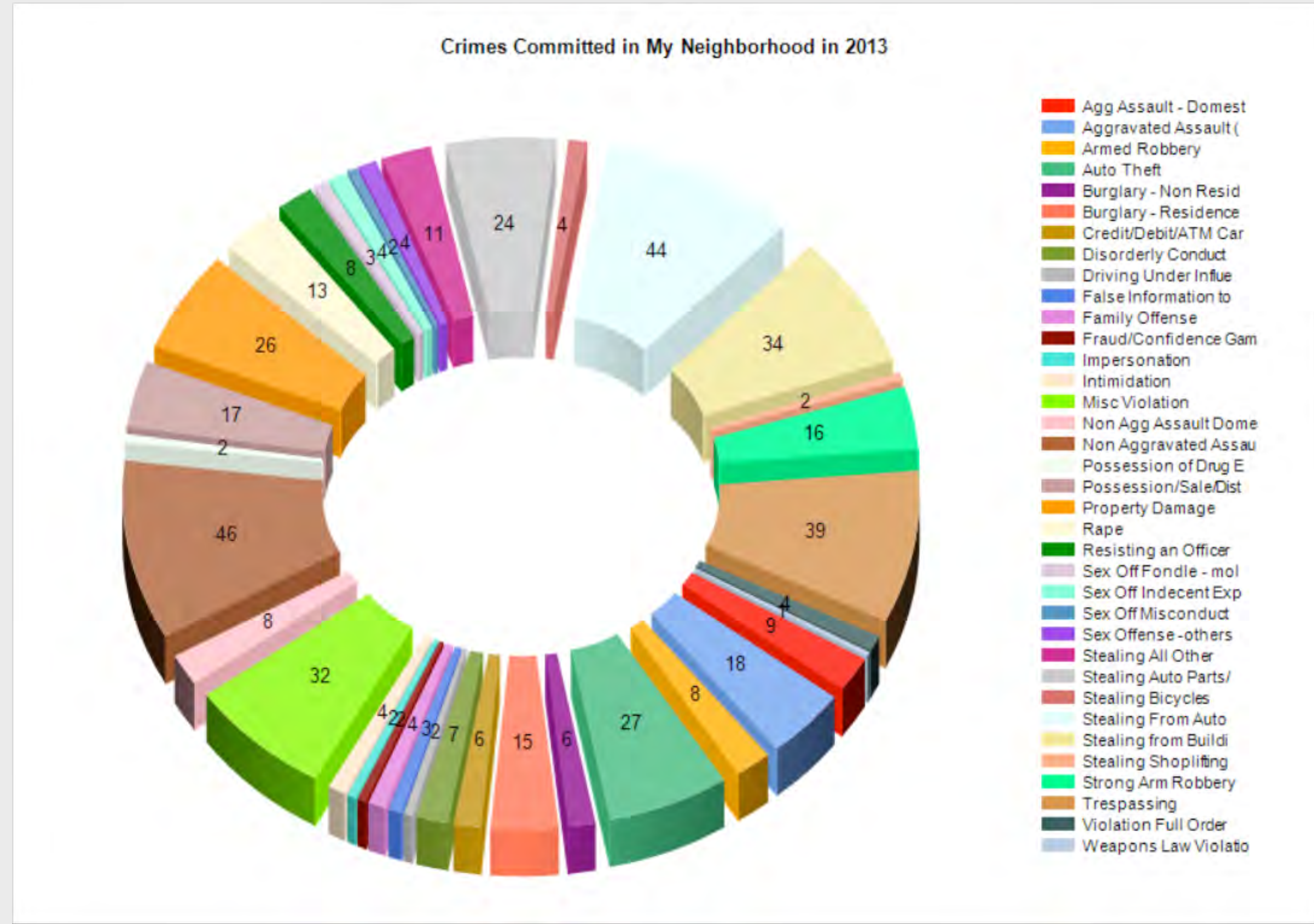
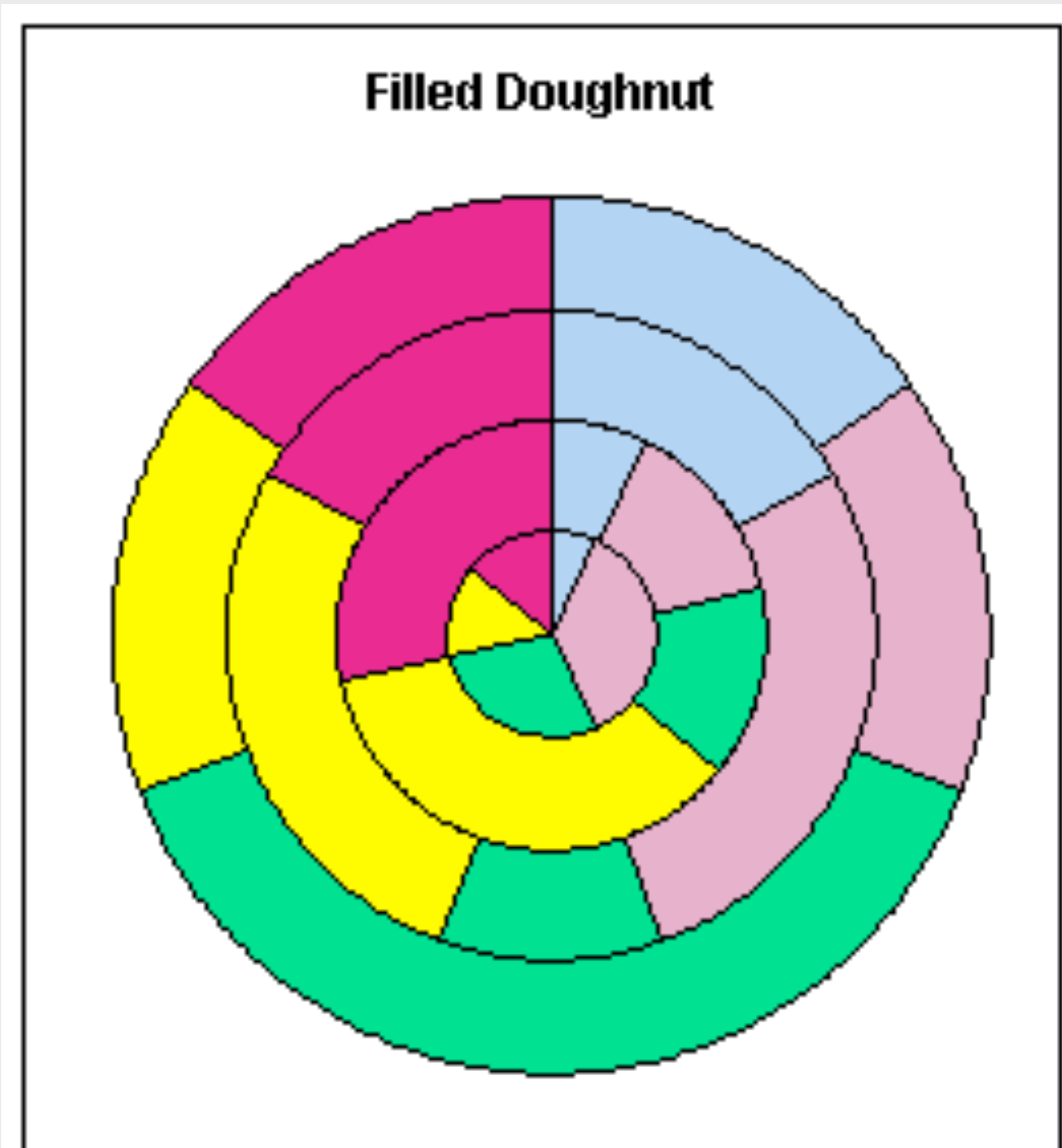
At least they didn't use a 3D pie chart like on the right



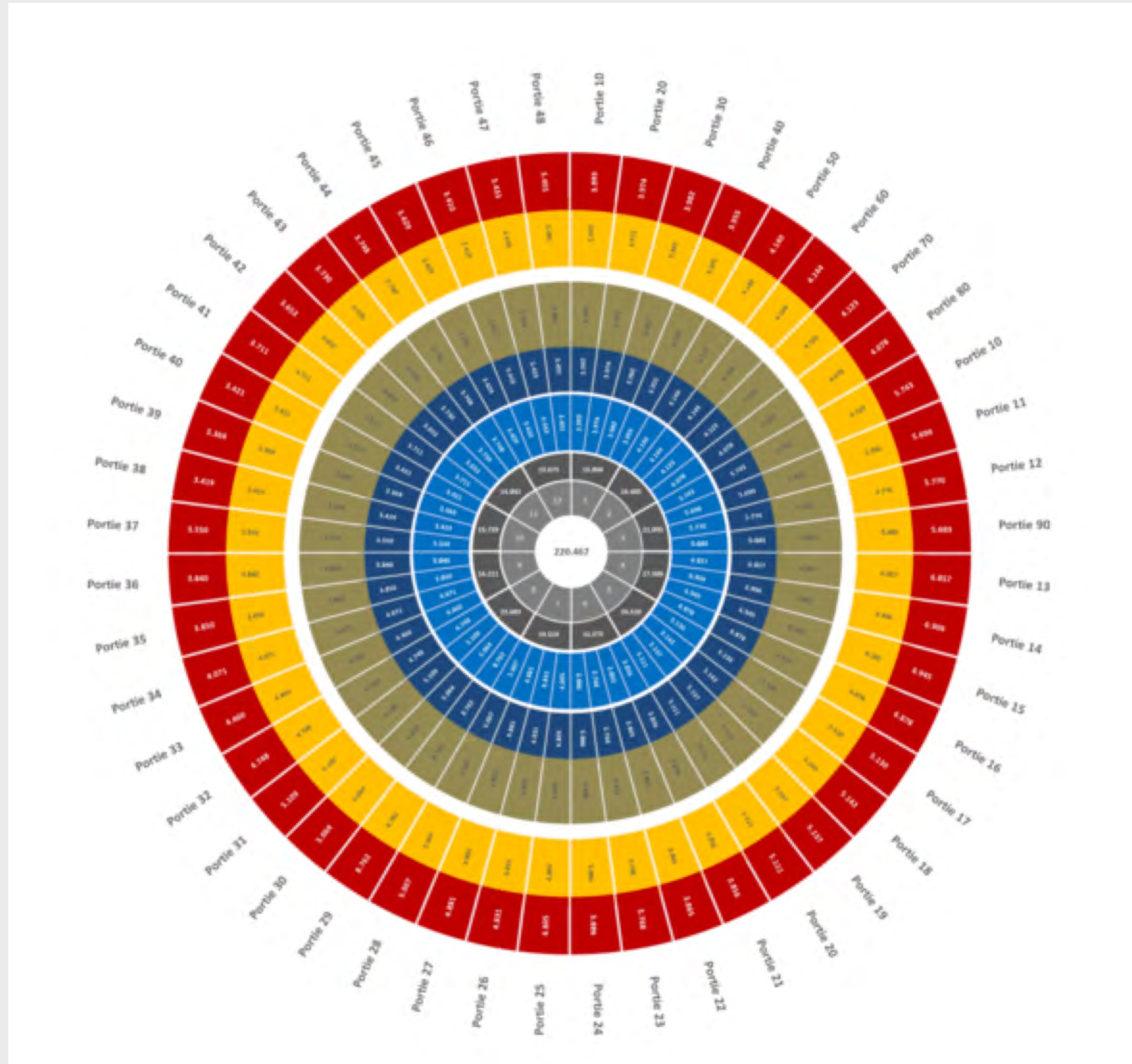


# Pie and Donut Charts

- Too many segments
- Not expressed as a percentage
- NO Cinnamon rolls

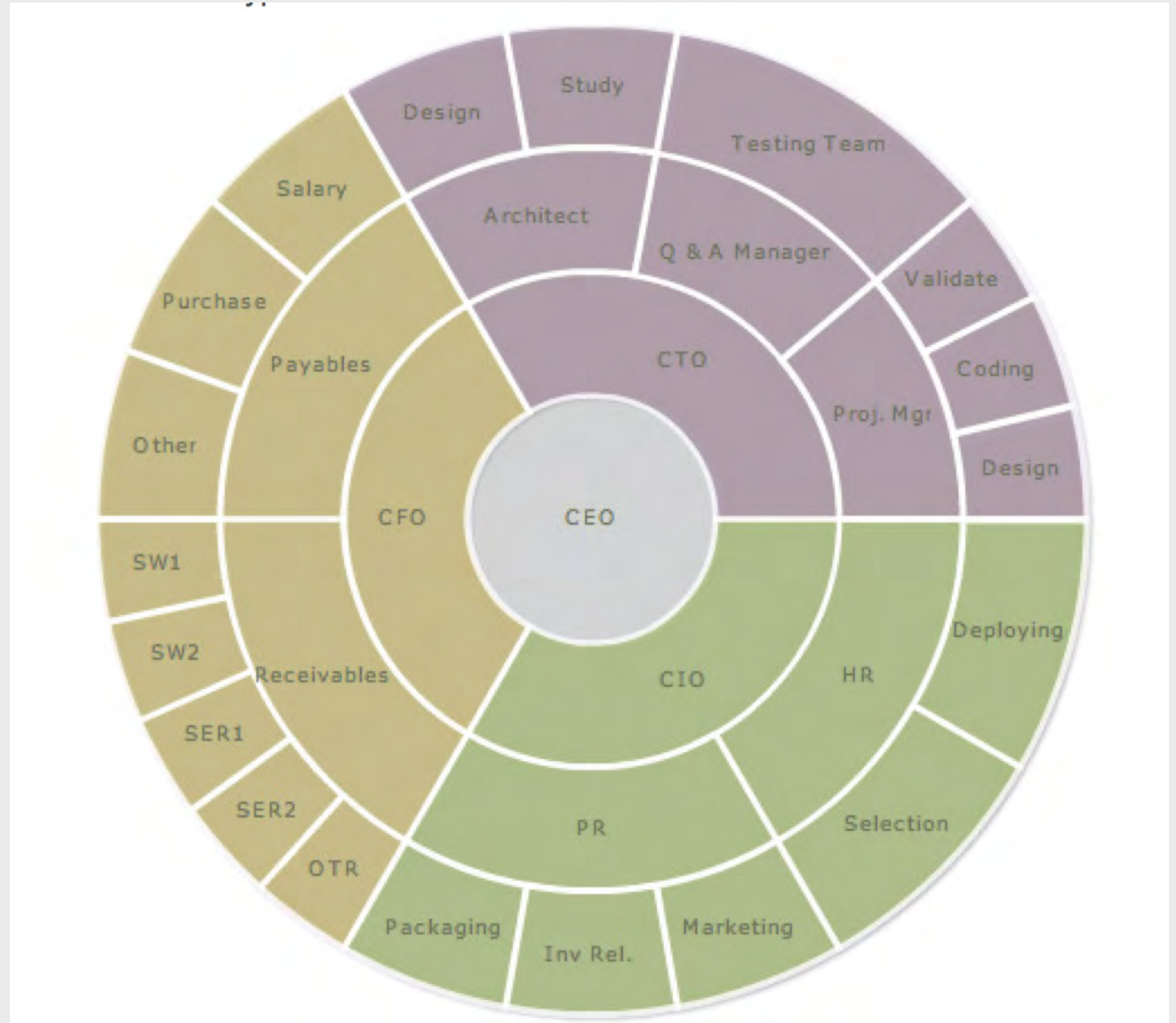


# Filled Doughnut #1



## Filled Doughnut #2

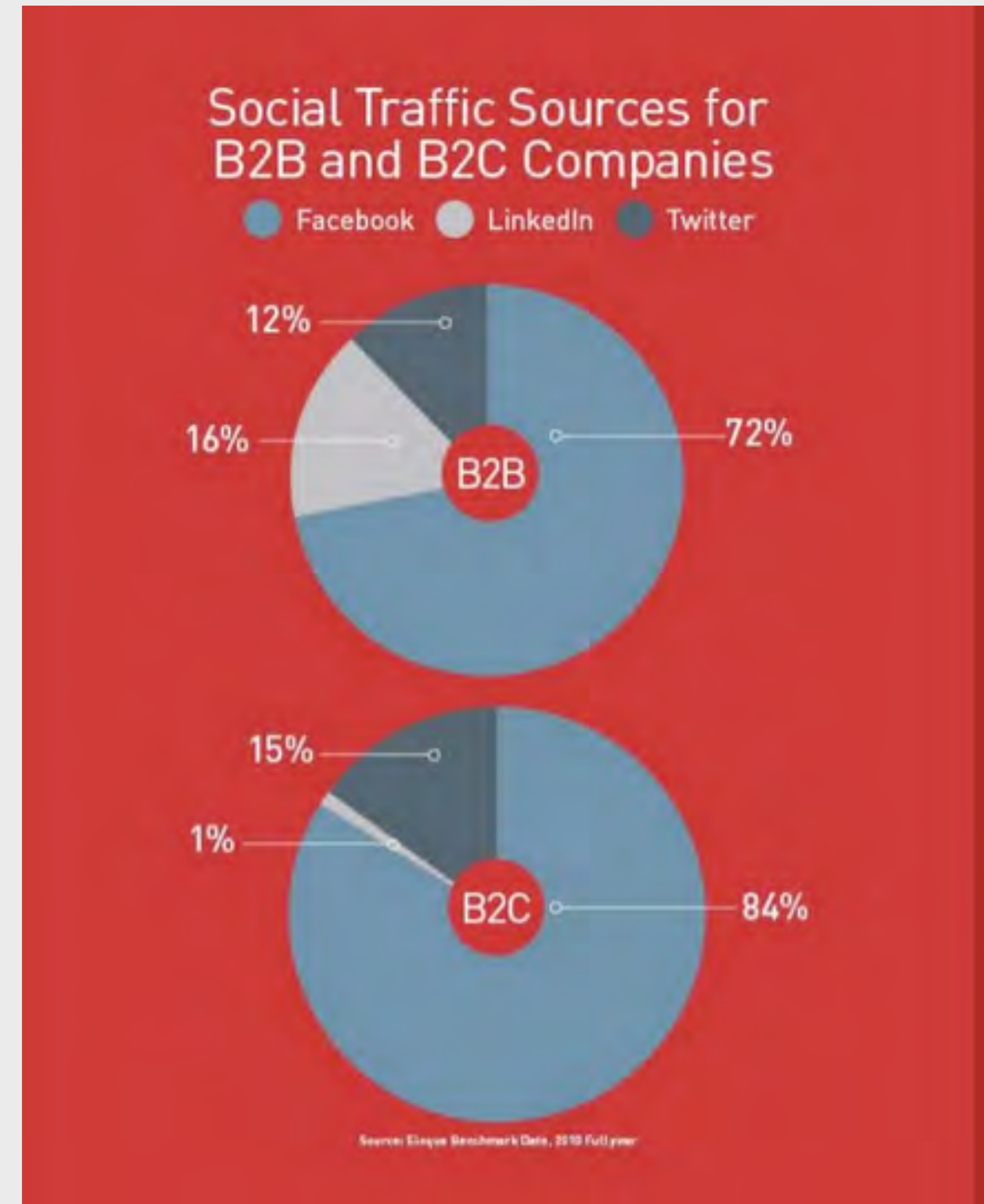
Example of a filled donut that makes sense. A tree diagram would work as well.



# Pie and Donut Charts

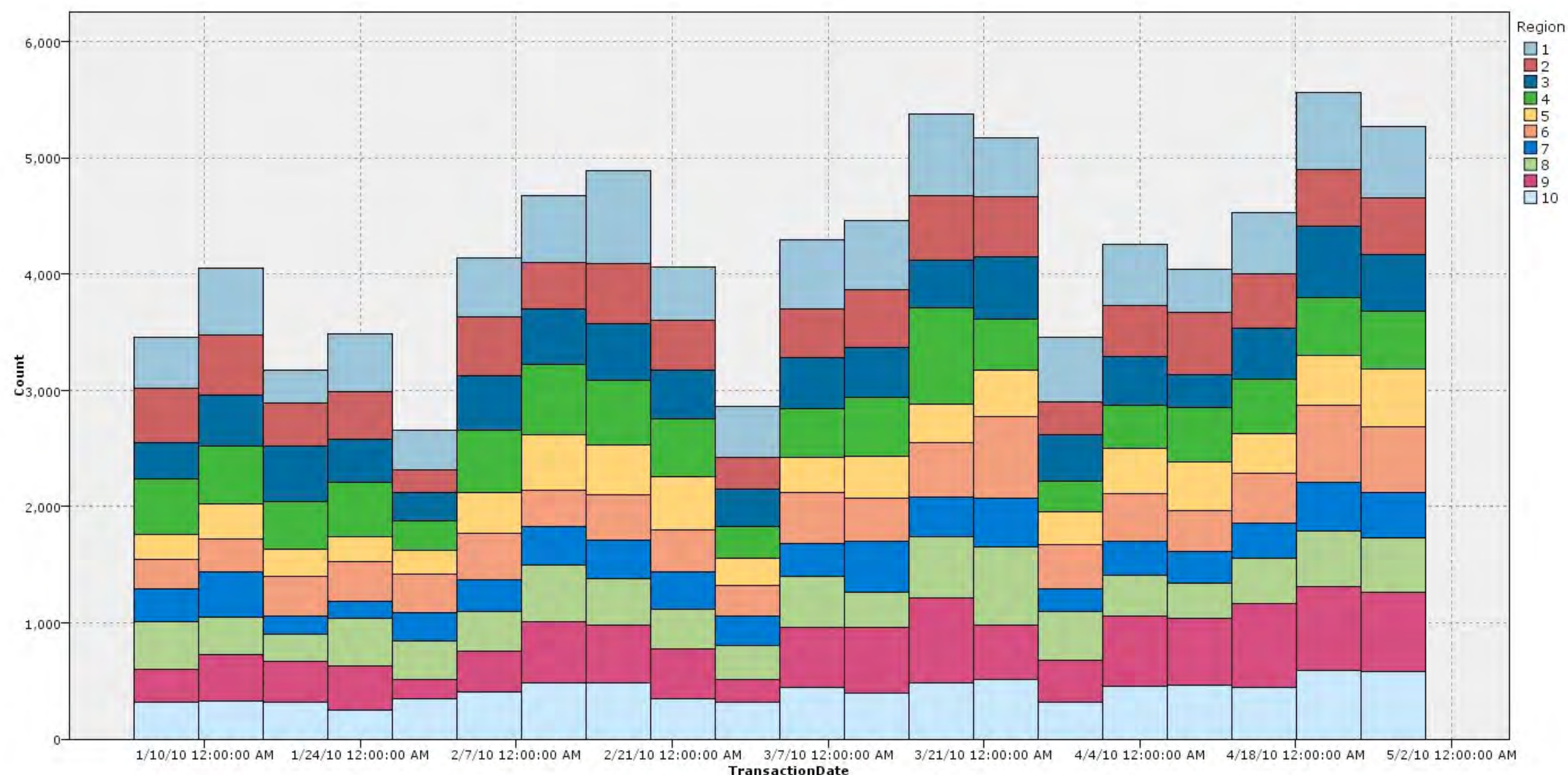
## Discussion:

Decompose the chart on the right and talk about what format would be best.



## Stacked Bar Graphs

- Are basically multiple pie charts shown in bar format.
- Only appropriate when the parts for each item in the series adds up to 100%
- Items in the Series are categorical (tasks, participants)
- Use color coding conventions your audience will understand
- The more bars, the more difficult to interpret



### Discussion:

What is **WRONG** with the chart on the left?

## Summary

- Knowing your Data is critical
  - The specific type of data you have will dictate the types of statistics you can use (or can't use)
- When presenting your data graphically, use the appropriate types of graphs.
  - Use Bar graphs for categorical data
  - Use line graphs for continuous data
  - Use pie charts or stacked bar graphs when sum to 100%
- Make them Attractive, yet MEANINGFUL and easy to Comprehend

*and.....*



# Usability Testing

August 2014



# Types of Usability Tests

## Formative Testing

- Is Creative in Nature
- Is ITERATIVE
- Can require hard data, but mostly provides observations
- Requires fewer users
- Conducted during development
- Used to get feedback, ideas and/or to determine if a specific design objective is being met

## Questions Answered

- What are the most significant issues at this time?
- What aspects work well and which do not?
- What are the most common errors being made?
- Are improvements being made in-between designs?
- What issues can be expected to remain after launch?

## Summative Testing

- Is Diagnostic in Nature
- Occurs upon Release
- Requires hard data (TOT, Error Rate, % completion)
- May require a large sample size (to achieve statistical significance)
- Conducted at the end of development
- Used to VALIDATE that a product has accomplished intended goals

## Questions Answered

- Have the usability goals been met?
- What is the overall usability?
- How does the product compare to the competition?
- Have improvements been made since last release?

# Types of Usability Testing Approaches

## Small Scale Qualitative

- Observation based
- Based on expertise and judgement of the specialist
- Specialist observers, interprets, decides

## Important Points

- Finding Issues (to be fixed)
- NOT finding statistics about your users
- NOT for measurement, comparison, generalization
- NOT a beginner's technique
- NOT UT-Lite

The CORRECT technique to use when iterating your Design

## Large Scale Quantitative

- Focused on quantitative measures
- Requires large sample size
- Results can be generalized to a larger population

## Important Points

- For finding statistics about your users
- For measurement, comparison, generalization
- Can include qualitative measures
- Formal Lab or remote testing

The CORRECT technique to use when evaluating your Design

# Small Scale Qualitative Testing

## Why?

- Can you afford to run a test with 300 users? - NO
- Can you afford to NOT do usability testing? - NO

## Correct Usage

- Ask the users to perform realistic tasks
- Use good sampling techniques
- Don't take quantitative measures - just look for UI design issues
- When you present results:
  - Don't report statistics, success rates, comparison
  - Present observations in this format
    - "Several users chose incorrectly, or seemed confused"
  - If allowed, present your solution

## Beware

- You can't generalize these results to the whole user group
- This is an exploratory, observational study, not a statistical analysis

# Small Scale Qualitative Testing

## Why it's so good

- Test once with X users - find up to 95% of the issues
- Redesign to fix the issues
- Test again with X users - find 95% of the issues
- Redesign to fix the issues
- Test a 3rd time with X users - find 95% of the issues

**Gives better cost/benefit than one test with 3X users.**

## But....

- Virzi, 1992 & Nielsen 1993 - 5 users will uncover 80% of usability problems
- Spool & Schroeder 2001 - 5 users will uncover 35% of the usability problems
- Faulkner 2003 - 5 users will uncover 85% of the usability problems, however
  - Percentage RANGED from nearly 100% to 55%
  - 95% of issues can be detected reliably with 10 users w/a lower bound of 82%

**Discussion: What is a potential problem with this?**

# Small Scale Qualitative Testing

## Nielsen's "5" users ...

"Why did you use only 5 people?"

- "Because that's what Nielsen says to use"

# WRONG

The Nielsen rule says that:

$$N(1-(1-L)^n)$$

Where **N** is the total number of usability problems in the design (and if we knew that we'd know what they are and would go ahead and fix them!)

Where **L** is the proportion of usability problems discovered while testing a single user (which you can only calculate AFTER you know N, which you can't up front, and after you have run your tests with that/those users)

THEIR typical value based on THEIR research is 31%

# *Small Scale Qualitative Testing*

Base the number of users on

- The criticality of the project (> Criticality > # of users)
- The experience of the testing team (> experienced team < # of users)
- The expected variety of the users (> diverse population > # of users)
- The complexity of the interface (> complexity > # of users)
- Whether the test is remote or on-site (remote is generally cheaper so you can do more)
- Your budget (> \$ > # of users)

## Small Scale Qualitative Testing

*"Although practitioners like simple directive answers such as the 5-user assumption, the only clear answer to valid usability testing is that the test users must be representative of the target population. The important and often complex issue, then, becomes defining the target population. There are strategies that a practitioner can employ to attain a higher accuracy rate in usability testing. One would be to focus testing on users with goals and abilities representative of the expected user population. When fielding a product to a general population, one should run as many users of varying experience levels and abilities as possible. Designing for a diverse user population and testing usability are complex tasks. It is advisable to run the maximum number of participants that schedules, budgets, and availability allow. The mathematical benefits of adding test users should be cited. More test users means greater confidence that the problems that need to be fixed will be found."*

# Small Scale Qualitative Testing

## General Rule


- More participants will generate more reliable results
- More participants generate fewer new results
- Test 10 participants per unique user group

**Hey! User Groups are IMPORTANT!**

**It is Important to Recruit from the Target User Group**

**ATM for an Indian market**

- Common icon for cash: hand-holding green bills
  - Doesn't communicate – Indian money is not green
  - Rupees are yellow, pink or light blue



The image shows a screenshot of an ATM screen with a transaction menu. A yellow circle highlights the 'Cash' icon, which depicts a hand holding green bills. An inset image shows a hand holding green bills, and another inset shows Indian Rupee banknotes. The ATM screen also displays the text 'Please select your transaction' and 'you will receive'.



# Quantitative Testing

## General Rule

- For Pure Measurement
- Performance: Time on Task, Error Rates, % Complete, keystrokes
- Satisfaction: Changes in subjective measurements
- Must be a benchmark or a repeat to compare scores
- The number of users is determined by the experimental design and the statistics used for analysis
- ...and yes, you need to have Hypotheses

### **Do Performance and Satisfaction Measures Always Correlate?**

*NO - We've repeatedly seen users who struggle give glowing ratings and have seen users that have had no issues give bad reviews. (Moderator Bias? Personal Bias? Badly designed test?)*

*In general,*

*the longer it takes for a user to complete a task, the lower the satisfaction as the completion rate increases (ability to do the task), the higher the satisfaction*

*\*\* Caveat: May NOT be valid for 'browse' and 'recreational' websites*

# Quantitative Testing

How to.....

- Set the usability criteria early (can be simple or complex)
- Focus on measurable behaviors
  - What are your metrics?
- Use Scenarios and Tasks
  - Tasks - something you tell the user to do
    - Quick and to the point
  - Scenarios - a general situation (story)
    - Some people consider a task a subset w/i a scenario
    - Provide motivation, rationale and context
    - Set the stage or tell a story
    - Make the situation more realistic
- Decide upon your experimental method
- Decide upon statistics being used

# Quantitative Testing

## Metrics

Usability Study Scenario	Task Success	Task Time	Errors	Efficiency	Learn-ability	Issues-based Metrics	Self-reported Metrics	Behavioral & Physiological Metrics	Combined & Comparative Metrics	Live Website Metrics	Card-Sorting Data
1. Completing a transaction	X			X		X	X			X	
2. Comparing products	X			X			X		X		
3. Evaluating frequent use of the same product	X	X		X	X		X				
4. Evaluating navigation and/or information architecture	X		X	X							X
5. Increasing awareness							X	X		X	
6. Problem discovery						X	X				
7. Maximizing usability for a critical product	X		X	X							
8. Creating an overall positive user experience							X	X			
9. Evaluating the impact of subtle changes										X	
10. Comparing alternative designs	X	X				X	X		X		

Task Efficiency = Time on Task to completion

## Quantitative Testing - SUS Scale

The System Usability Scale (SUS) was released into this world by John Brooke in 1986.

It is a subjective measure of learnability, usability and satisfaction

SUS is technology independent and has since been tested on hardware, consumer software, websites, cell-phones, IVRs and even the yellow-pages.

It has become an industry standard with references in over 600 publications and correlations are consistent across over 1500 studies.

It can be used in both formative and evaluative research

It is a 5 point scale (don't change it)

It is scored a strangely

- For odd numbered items: subtract one from the user response

- For even-numbered items: subtract the user responses from 5

- This scales all values for 0 to 4 (with 4 being the most positive response)

- Add up the converted responses for each user and multiply that total by 2.5.

  - This converts the range from 1 to 100 instead of from 0 to 40

# Quantitative Testing

## SUS Scale

1. I think that I would like to use this system frequently

Strongly disagree					Strongly agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
1	2	3	4	5	

2. I found the system unnecessarily complex

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5

3. I thought the system was easy to use

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5

4. I think that I would need the support of a technical person to be able to use this system

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5

5. I found the various functions in this system were well integrated

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5

# Quantitative Testing

## SUS Scale

6. I thought there was too much inconsistency in this system

1	2	3	4	5

7. I would imagine that most people would learn to use this system very quickly

1	2	3	4	5

8. I found the system very cumbersome to use

1	2	3	4	5

9. I felt very confident using the system

1	2	3	4	5

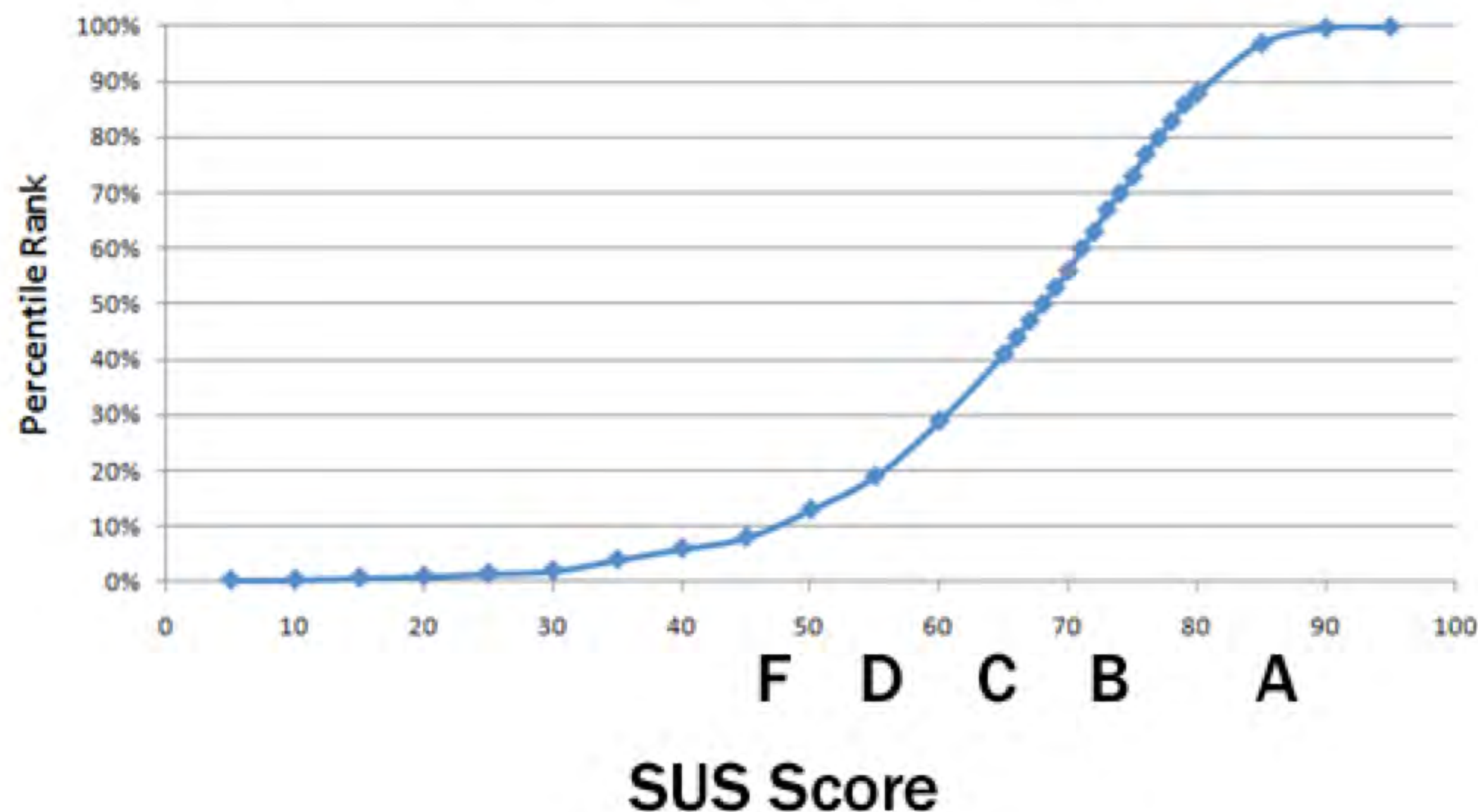
10. I needed to learn a lot of things before I could get going with this system

1	2	3	4	5

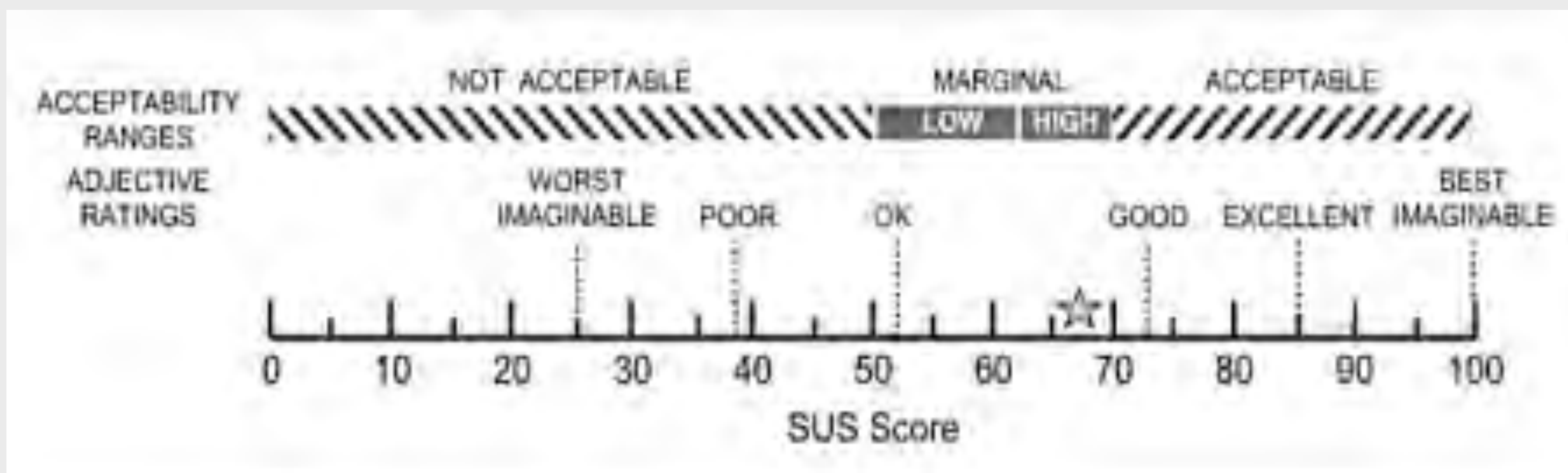
# Quantitative Testing

## SUS Scale What does it mean.....

The graph below shows how the percentile ranks associate with SUS scores and letter grades.



The average SUS score for a collection of 500 studies is 68



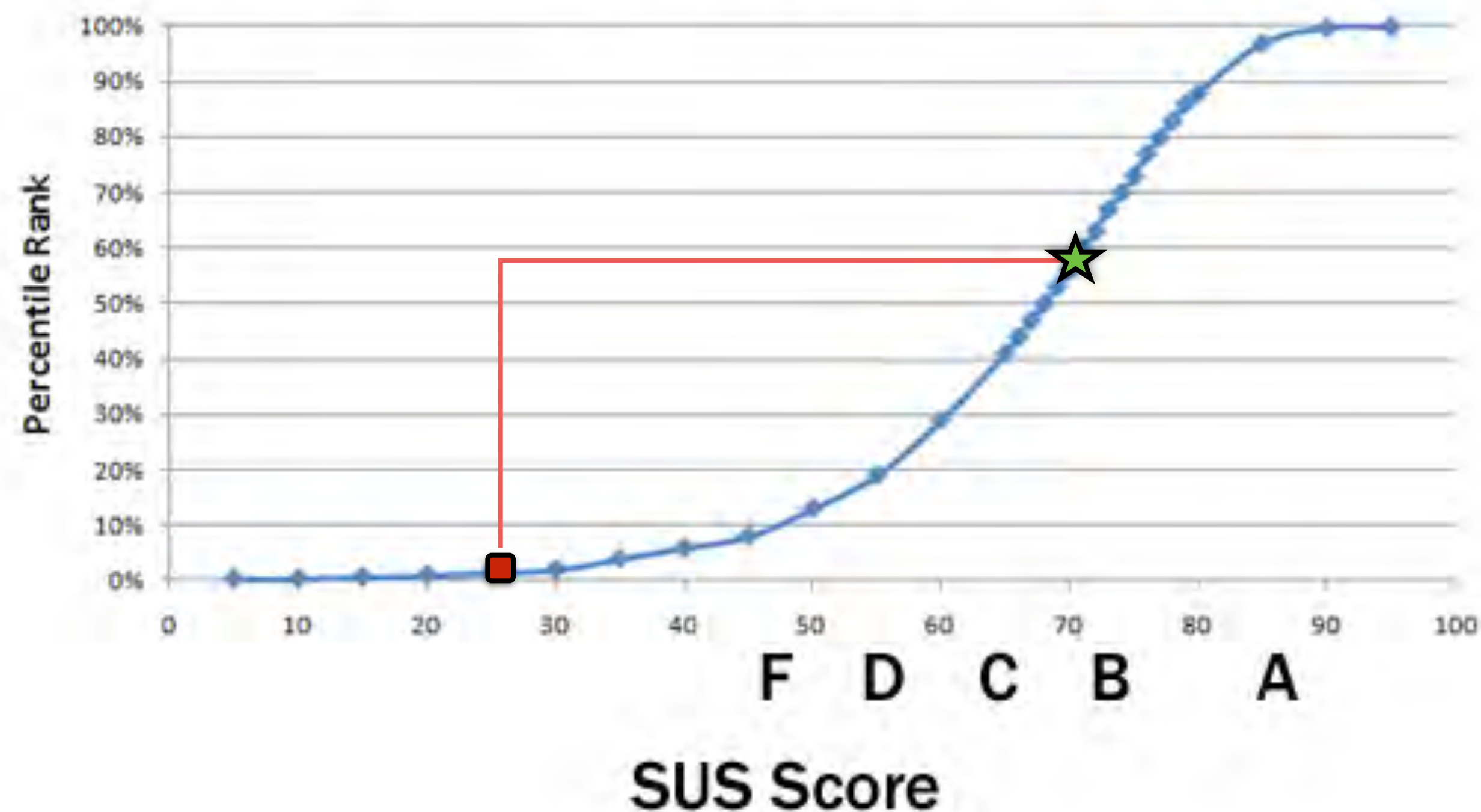
# Quantitative Testing

SUS Scale A recently tested interface

## Website Usability Scale (SUS) Rating - a measure of Usability, Learnability and user Satisfaction

- The Website Usability Scale Score was **25 out of 100** (while the average website SUS score is 68)

The graph below shows how the percentile ranks associate with SUS scores and letter grades.



SUS Rating Scale Items:

1. I think that I would like to use this website frequently.
2. I found the website unnecessarily complex.
3. I thought the website was easy to use.
4. I think that I would need the support of a technical person to be able to use this website.
5. I found the various functions in this website were well integrated.
6. I thought there was too much inconsistency in this website.
7. I would imagine that most people would learn to use this website very quickly.
8. I found the website very cumbersome to use.
9. I felt very confident using the website.
10. I needed to learn a lot of things before I could get going with this website.



## Summary

- Formative testing is creative and iterative and takes place during development; it focuses on observations
  - Test with a minimum of 10 participants per user group (depending on prior factors discussed)
  - Avoid moderator bias
  - Be wary of subjective ratings
- Summative testing is diagnostic in nature, focuses on performance metrics, and is used after development
  - Use the number of users required by your statistical method
  - Avoid moderator bias
  - Use the SUS scale for subjective Assessment
- When planning, consider whether to use tasks, scenarios, or self-guided evaluations

***SKIT***

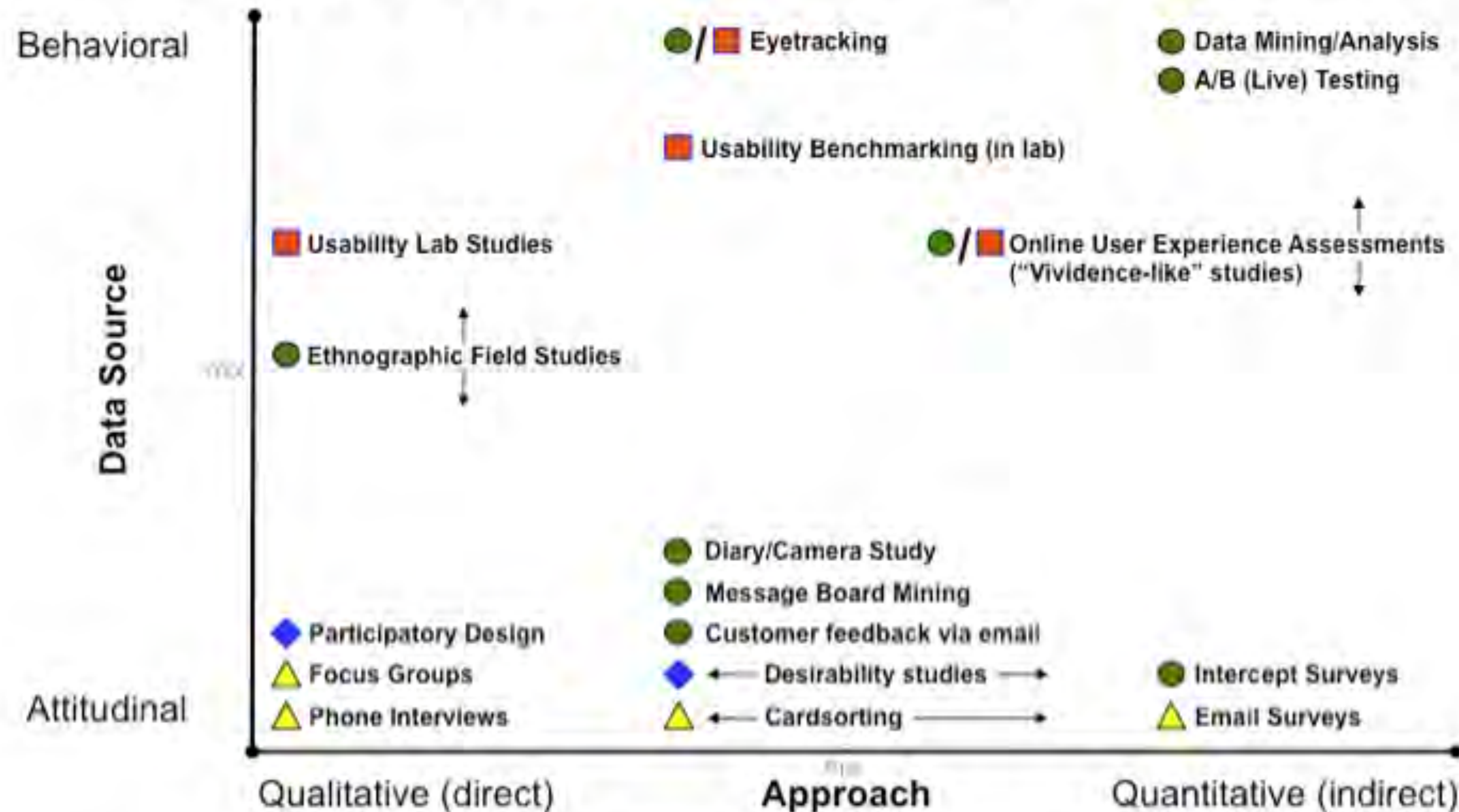
# Usability Methodology

August 2014

# Stuff we can do.....

The following chart illustrates where several popular methods appear along these dimensions:

Research methods by **Data Source** vs. **Approach** vs. **Context of Product Use**



**Key for Context of Product Use during data collection**

- Natural use of product
- ▲ De-contextualized / not using product
- Scripted (often lab-based) use of product
- ◆ Combination / hybrid

© 2008 Christian Rohrer

Don't get Qualitative and Quantitative on the horizontal axis confused with Qual and Quan data; it is simply stating whether the measurements are DIRECT or INDIRECT w/respect to the users

These are research methods. Some are NOT usability Tests This is NOT a complete list

Focus Groups are NOT usability tests; they are subject to extreme bias in their natural form.

Each dimension provides a way to distinguish between studies in terms of the questions they answer and the kinds of purposes they are most suited for.

# When can we do stuff ....

Planning & Feasibility	Requirements	Design	Implementation	Test & Measure	Post Release
Getting started	User Surveys	Design guidelines	Style guides	Diagnostic evaluation	Post release testing
Stakeholder meeting	Interviews	Paper prototyping	Rapid prototyping	Performance testing	Subjective assessment
Analyse context	Contextual inquiry	Heuristic evaluation		Subjective evaluation	User surveys
ISO 13407	User Observation	Parallel design		Heuristic evaluation	Remote evaluation
Planning	Context	Storyboarding		Critical Incidence Technique	
Competitor Analysis	Focus Groups	Evaluate prototype		Pleasure	
	Brainstorming	Wizard of Oz			
	Evaluating existing systems	Interface design patterns			
	Card Sorting				
	Affinity diagramming				
	Scenarios of use				
	Task Anaysis				
	Requirements meeting				

# RITE Testing

August 2014

# Rapid Iterative Test and Evaluation Method (RITE)

## History & Definition

- RITE methodology was originally identified through the observation of a QA technique employed by the Microsoft gaming division
- It was later been adapted to usability testing in the early 00's.<sup>1</sup>
- It is believed that it might overcome the barrier created by time and monetary constraints that come with thorough user testing and multiple iterations.
- As a usability testing methodology, RITE is a 'discount' usability test that is conducted in a fast and highly collaborative manner. <sup>2</sup> (see Figure 1) It's ONLY purpose is to identify and fix as many issues as possible and to verify the effectiveness of those fixes in the shortest possible time.

It has been shown to be effective in the following *specific contexts*:

- electronic real-time game development
- emerging software technologies, where there are more problems to be solved than traditional usability methods can handle (many surface level, easily discoverable issues)
- where rapid growth and change is compressing software release cycles.

<sup>1</sup> Medlock, MC, Wixon, D, Terrano M, Romero R, & Fulton B (2002). Using the RITE Method to improve product: a definition and a case study. UPA2002, Orlando, FL, July 2002

<sup>2</sup> Bias, RG and Mayhew, JD (2005) Cost Justifying Usability: and Update for an Internet age, Elsevier.

## Rapid Iterative Test and Evaluation Method (RITE)

### Description

As figure 1 shows, the RITE method emphasizes rapid changes to the design as soon as a usability problem is identified and a solution is available (even after the feedback from a single participant)

The benefit of incorporating quick changes is that the next participant can evaluate the “improved” design, reducing the design process time. As implied by figure 1, RITE is comparable with Agile methodologies due to its iterative nature.

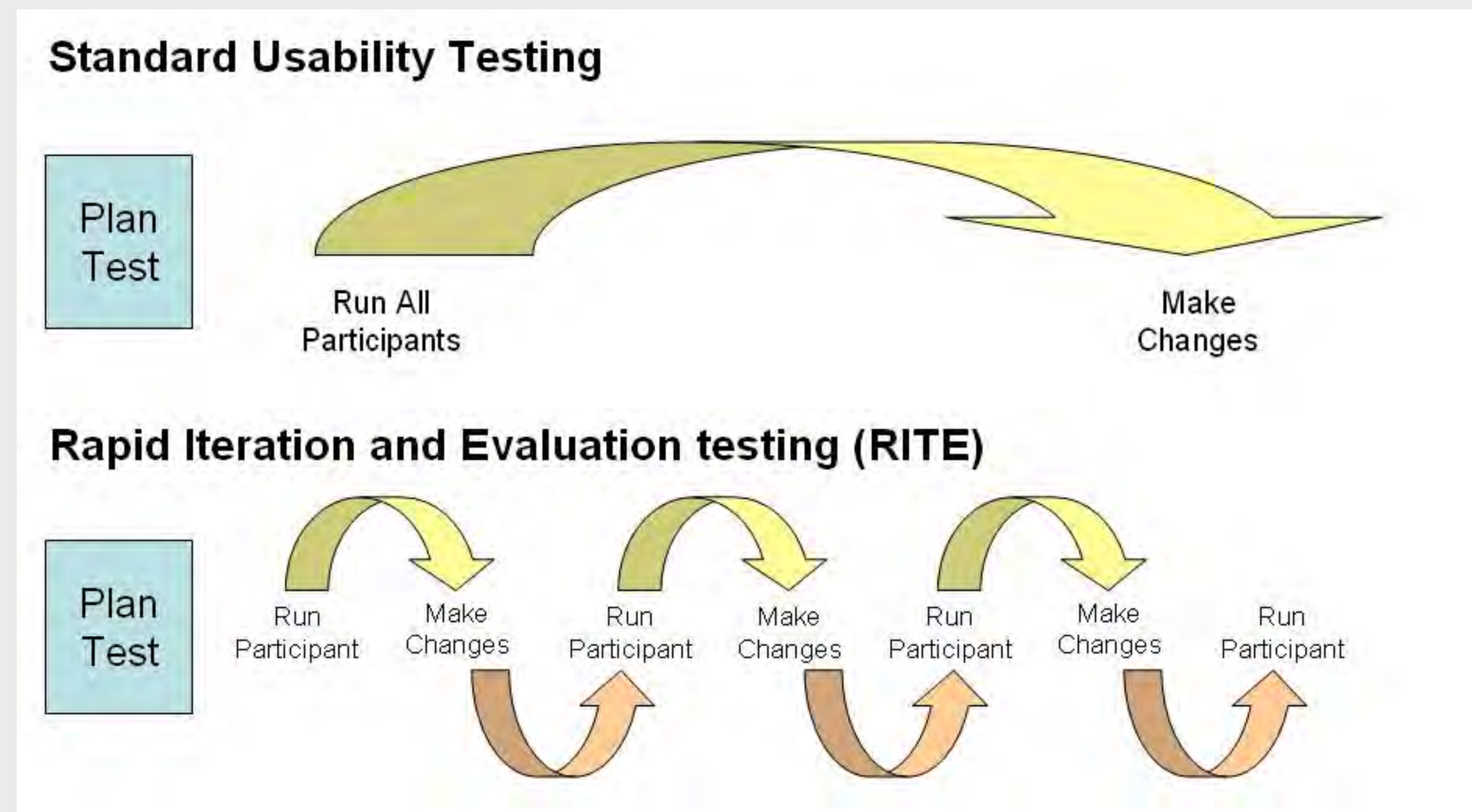


Figure 1. RITE testing versus “Standard” Usability Testing <sup>1</sup>

<sup>1</sup> Salvendy, G (2012), Handbook of Human Factors and Ergonomics, Fourth Edition, John Wiley & Sons, Inc.

<sup>2</sup> Sharritt, M., (2013) What’s wrong with the RITE method?



## Rapid Iterative Test and Evaluation Method (RITE)

### Description

- RITE is a tool for “designers” to use when they are under time constraints and should NOT replace traditional methods that are more thorough in nature, as those methods tend to provide the most complete usability findings. <sup>1</sup>
- RITE is ‘Formative’ in nature in that it impacts the design through iterative testing and rapid changes. It cannot be used in an ‘evaluative’ manner as it does not have any scientific rigor or statistical power.
- RITE observers are made up of designers, developers and those who have the ability to authorize changes for the interface. They are present *at all times*, so that they are ready to discuss and implement changes *immediately*, if not within 24 hours, of a single test session.
- RITE is specifically appropriate for finding issues with higher level task specific issues and NOT open-ended in-depth analysis which is the output of more traditional methods. <sup>2</sup>

<sup>1</sup> Salvendy, G (2012), Handbook of Human Factors and Ergonomics, Fourth Edition, John Wiley & Sons, Inc.

<sup>2</sup> Sharritt, M., (2013) What’s wrong with the RITE method?

## Rapid Iterative Test and Evaluation Method (RITE)

### Requirements <sup>1</sup>

- **Traditional Laboratory setup** with computers, screen captures software, audio/video recording and observation area (or remote viewing capability)
- **Test participant pool** that represents the target user population.
- **Usability engineer(s)** that have experience both in the intended software domain and in the problems that users typically experience in that domain. <sup>2</sup>
- **Key Decision Makers** (Those who have the authority to authorize a change to the user interface) must observe all the participants with the usability engineer(s) in real time.
  - The presence of these individuals is critical to the success of RITE, as their combined input and real-time collaboration is needed to analyze the impacts and feasibility of any potential changes.
  - This may include members of the UXD, VDT, Programming and Architectural teams.

<sup>1</sup> Bias, RG and Mayhew, JD (2005) Cost Justifying Usability: and Update for an Internet age, Elsevier.

<sup>2</sup> Medlock, MC, Wixon, D, Terrano M, Romero R, & Fulton B (2002). Using the RITE Method to improve product: a definition and a case study. UPA2002, Orlando, FL, July 2002

## Rapid Iterative Test and Evaluation Method (RITE)

### Requirements <sup>1</sup>

- **Data analysis** occurs after each participant, or at least after each day of testing.
- **Dedicated FED/BED developers** make changes as soon as a problem is identified and a potential solution is clear during the course of each testing session (no more than 24 hours lag time)
- **Modifications are tested** with subsequent users to see if the changes solved the uses previously identified without “devolving” the interface (interjecting new problems).

<sup>1</sup> Bias, RG and Mayhew, JD (2005) Cost Justifying Usability: and Update for an Internet age, Elsevier.

## Rapid Iterative Test and Evaluation Method (RITE)

### Strengths <sup>1</sup>

- **Decreased Costs** (if used *within appropriate contexts*)
  - Validated Contexts include: electronic real-time strategy games, emerging software technologies, where there are more problems to be solved than traditional usability methods can handle, and where rapid growth and change is shortening inflexible software release cycles.
- **Increased defect fix rate** (an effect of design/product teams 'watching' the testing)
  - Increases team members' willingness to make changes
- **Potential for more superficial issues to be found** (if discovered and fixed issues are 'blocking' in nature)
- Results in **better team dynamics** (if appropriate teams / team members are included)
- **Degrades gracefully** into a very basic (but very expensive) discount usability test if 'decision makers' or 'resources' back out
- **Integrates well (in theory) with Agile** methodologies of software development.

<sup>1</sup> Bias, RG and Mayhew, JD (2005) Cost Justifying Usability: and Update for an Internet age, Elsevier.

## Rapid Iterative Test and Evaluation Method (RITE)

### Weaknesses <sup>1</sup>

- **RITE is Context Sensitive** (what is being tested as well as too many or too few decision makers being involved influences the outcome)
- **RITE is sensitive to the number of changes** being made; too many at once may degrade usability while making one at a time may be cost prohibitive
- **RITE is follow-up sensitive** in that after the final changes are made, the interface must be tested with enough users to ensure that no issues have been missed or interjected into the design (5 users after the FINAL error and failure is observed).
  - With a poorly thought out test or highly complex interface this could increase the number of participants and become cost prohibitive
- **RITE is not suitable for discovering deeply-rooted issues or mission critical situations**, for its 'rapid iterative' approach fails to conduct an open-ended, in-depth analysis thus failing to identify serious infrequently occurring issues.<sup>2</sup>

<sup>1</sup>Bias, RG and Mayhew, JD (2005) Cost Justifying Usability: and Update for an Internet age, Elsevier.

<sup>2</sup> Sharritt, M., (2013) What's wrong with the RITE method?

## Rapid Iterative Test and Evaluation Method (RITE)

### Weaknesses <sup>1</sup>

- **RITE method can take up to over two weeks** to complete; this is due to the large number of participants needed and the time between sessions to make changes to the system.<sup>1</sup>
- **RITE may require large numbers of participants** if fixes are to be validated (which they should be).<sup>2</sup>
- **The end of RITE testing is hard to estimate** if you continue to test until all fixes have been validated.

<sup>1</sup>Bias, RG and Mayhew, JD (2005) Cost Justifying Usability: and Update for an Internet age, Elsevier.

<sup>2</sup> Medlock, M.C., Wixon, D., Terrano, M., Romero, R., & Fulton, B. (2002).Using the RITE method to improve products: A definition and a case study. Presented at the Usability Professionals Association, Orlando Florida.

## Rapid Iterative Test and Evaluation Method (RITE)

Where does RITE fit into the Usability Testing continuum? <sup>1</sup>

- RITE exists within the continuum of the *software design phase* as it is too advanced for requirements gathering and not appropriate for the test and measure phase of development. Its optimal usage is prior to implementation while there is still time to make design changes if it is used within the appropriate context. This statement assumes that all major 'meta' issues have been resolved by the time RITE is utilized.
- Given that each usability evaluation method has an optimal context in which to operate it is hard to define where it might fit within the continuum. Instead, the below table shows the appropriate methods for the design stage in increasing order complexity and cost.

SDLC Design Phase Activities (> cost & > complexity)
Paper (Lo-Fi) Prototyping (one pathway)
Simplified user testing (with prototype or paper multiple pathways)
Inspection Methods (Heuristic Evaluation)
RITE Method
Traditional laboratory Testing

## Rapid Iterative Test and Evaluation Method (RITE)

How does RITE compare to other methods?

Resource	RITE Usability Test	Traditional Usability Test	Concept Test	MiLE	NN Heuristic Eval
Lab Set Up	Full lab Set Up	Full Lab Set Up	Office, Computer, recording	None	None
Stimulus Materials	Hi-Fi Prototype	Hi-Fi Prototype	Wireframes, Lo-Med-Hi Fi prototype	Wireframes to working prototype	Wireframes to working prototype
Vetted Test participant Pool	Yes	Yes	Depends	Maybe	Expert reviewers only
Usability Engineers Types	SME	Generic	Generic	Generic	Generic
UXD Observers	Yes	Optional	Optional	No	No
VDT Observers	Yes	Optional	Optional	No	No
FED/BED Observers	Yes	Optional	Optional	No	No
UXA Observers	Yes	Optional	Optional	No	No
Changes Made	Every 24 Hours max.	At end of Test Run (1 to 2 weeks)	At end of test run	At end of Eval Run (1 to 3 days)	At end of Eval Run (1 to 3 days)
Results Analyzed	Per Participant	At end of Test Run (1 to 2 weeks)	At end of test run	At end of Test Run (1 to 3 days)	At end of Test Run (1 to 3 days)
Number of subjects	Number to get to 0 Defects / Errors + 5	Number to bring statistical significance per test condition	10 to 15	3 to 5	3 to 5
Finds Superficial Problems	Yes	Yes	yes	Yes	Yes
Finds 'Meta' / Deep / rarely occurring errors?	No	Yes	Maybe	no	No
Statistical Significance?	No	Yes	No	No	No
Time to Test subjects	2 weeks	2 weeks	2-3 days	2 days	2 days
Report Generation?	No	Yes	Yes	Yes	Yes



# Rapid Iterative Test and Evaluation Method (RITE)

## Summary

- **RITE is Context Sensitive** (what is being tested as well as too many or too few decision makers being involved influences the outcome)
- **RITE is sensitive to the number of changes** being made; too many at once may degrade usability while making one at a time may be cost prohibitive
- **RITE is follow-up sensitive** in that after the final changes are made, the interface must be tested with enough users to ensure that no issues have been missed or interjected into the design (5 users after the FINAL error and failure is observed).
  - With a poorly thought out test or highly complex interface this could increase the number of participants and become cost prohibitive
- **RITE is not suitable for discovering deeply-rooted issues or mission critical situations**, for its 'rapid iterative' approach fails to conduct an open-ended, in-depth analysis thus failing to identify serious infrequently occurring issues.<sup>2</sup>

<sup>1</sup>Bias, RG and Mayhew, JD (2005) Cost Justifying Usability: and Update for an Internet age, Elsevier.

<sup>2</sup> Sharritt, M., (2013) What's wrong with the RITE method?

# Card Sorting

August 2014

# Card Sorting

Instead of organizing a website by some byzantine corporate structure, you base it on how the users think by having them sort items into categories.

Useful approach for designing information architecture, workflows, menu structure, or web site navigation paths.

Variations of Card Sorting have been used in psychological research for around 100 years

Card sorting is applied when

- the variety in the items to be organized is so great that no existing taxonomy is accepted as organizing the items.
- The similarities among the items make them difficult to divide clearly into categories.
- Members of the audience that uses the environment may differ significantly in how they view the similarities among items and the appropriate groupings of items.

# Card Sorting

(open)

## Definition

In an open card sort, participants create their own names for the categories.

-This helps reveal not only how they mentally classify the cards, but also what terms they use for the categories.

-Open sorting is **generative**; it is typically used to discover patterns in how participants classify, which in turn helps generate ideas for organizing information.

## Operation:

-A person representative of the audience is given a set of index cards with terms already written on them.

-Variation: Use Terms and their definitions along with common language and no acronyms

-The person puts the terms into logical groupings, and finds a category name for each grouping.

-Variation: After each grouping session, ask participant to re-evaluate and see if any groups could contain other groups identified.

-This process is repeated across a population of test subjects

-The results are later analyzed to reveal patterns.

-Use a software package; do not try to do this manually at home (can be hazardous to your mental health)

# Card Sorting

(Closed)

## Definition

In an closed card sort, participants have to sort into provided categories.

- This helps reveal how they mentally classify the cards in accordance with a previously delineated menu structure
- This helps reveal the degree to which the participants agree on which cards belong under each category.
- Closed sorting is **mostly evaluative**; it is typically used to judge whether a given set of category names provides an effective way to organize a given collection of content.

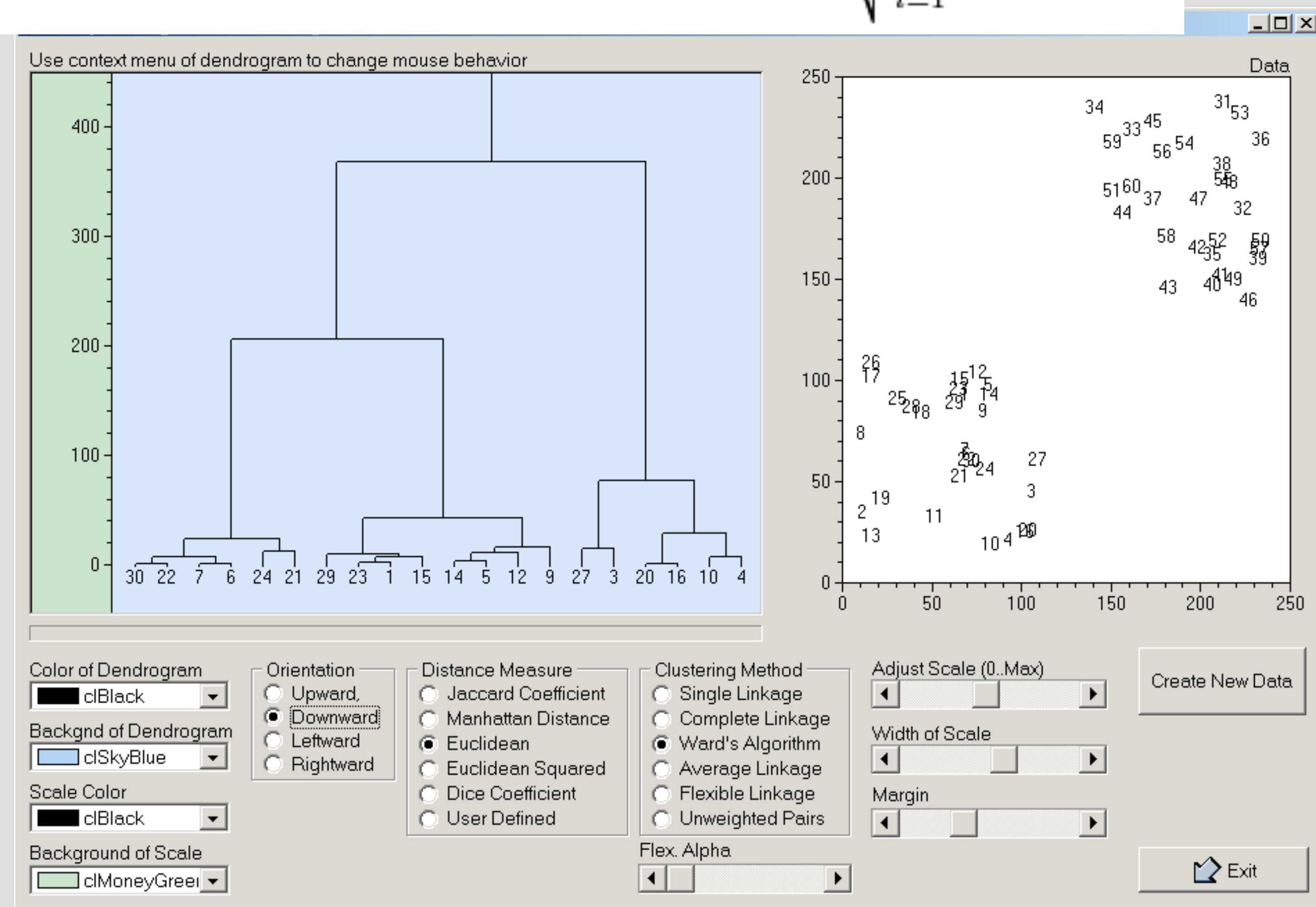
## Operation:

- A person representative of the audience is given a set of index cards with terms already written on them.
  - Variation: Use Terms and their definitions along with common language and no acronyms
- The person puts the terms into logical groupings, and finds a category name for each grouping.
  - Variation: After each grouping session, ask participant to re-evaluate and see if any groups could contain other groups identified.
- This process is repeated across a population of test subjects
- The results are later analyzed to reveal patterns.
  - Use a software package; do not try to do this manually at home (can be hazardous to your mental health)

# Card Sorting Results

- Dendrograms and Clusters based on Euclidean Geometry

$$d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}.$$



Repeat, do NOT try this at home

# Tree Test

## Definition

- Tree testing is a usability technique for **evaluating** the **findability** of topics in a website. It is also known as reverse card sorting or card-based classification.
- A large website is typically organized into a hierarchy (a "tree") of topics and subtopics. Tree testing provides a way to measure how well users can find items in this hierarchy.
- Tree testing is not done on the website itself; instead, a simplified text version of the site structure is used. This ensures that the structure is evaluated in isolation, nullifying the effects of navigational aids, visual design, and other factors.

## Operation:

- The participant is given a "find it" task (e.g., "Look for brown belts under \$25").
- They are shown a text list of the top-level topics of the website.
- They choose a heading, and are then shown a list of subtopics.
- They continue choosing (moving down through the tree) — drilling down, backtracking if necessary — until they find a topic that satisfies the task (or until they give up).
- The participant does several tasks in this manner, starting each task back at the top of the tree.
- Once several participants have completed the test, the results are analyzed for each task.
- Analysis is somewhat easier to do by hand

# Review

- Is a Card sort evaluative or generative?
- When do you use a 'Tree Test' and what does it tell you?
- When diagnosing and fixing a current navigation structure what is the most prudent process in relation to what we talked about last time?
- Is it easy to analyze a tree test?
- What is something do you not want to do at home w/o software?
- When performing a tree test, do you tell the participants when they get it right or wrong?
  - Why?
- What are the two types of card sort tests?
- Should YOU the UX person or SME be a test participant in your own card sort/tree test?



# Review

- Is a Card sort evaluative or generative? - **Generative (building UP from APIs to pages and from pages to categories and from categories to relationships)**
- When do you use a 'Tree Test' and what does it tell you? - **When you need to evaluate an existing navigation structure; it tells you if your navigation SUCKS (87% correct is the cut off)**
- When diagnosing and fixing a current navigation structure what is the most prudent process in relation to what we talked about last time? -**Tree test, Card sort, Tree Test**
- Is it easy to analyze a tree test? -**Yes, it is simply %'s of correct choices with a cut off of acceptable at 87% correct**
- What is something do you not want to do at home w/o software? - **Analyze a Card Sort**
- When performing a tree test, do you tell the participants when they get it right or wrong? - **No**
  - Why? -**They will learn patterns and the purpose of the test is judge intuitive fundability**
- What are the two types of card sort tests? - **Open (user generates grouping names/categories) and Closed (user sorts into pre determined categories)**
- Should YOU the UX person or SME be a test participant in your own card sort/tree test?- **No your experience and domain knowledge invalidates you.**

# Heuristic Review

UserFocus

August 2014

# Heuristic Review

The main goal of Expert Reviews is to identify any problems associated with the design of user interfaces. Usability 'Experts' are used to review the interface instead of non-experts.

It is used EARLY in the development process during the design stage to catch potential issues BEFORE going to code.

Expert Reviews may or may not have a predetermined checklist depending on the expertise of the Reviewers, however...

We suggest using UserFocus' Heuristics for Expert Review (2009); 247 Checkpoints (<http://www.userfocus.co.uk/resources/guidelines.html>)

# Expert Review / Evaluation

## Definition:

An Expert evaluation is a usability inspection method for computer software that helps to identify usability problems in the user interface (UI) design. It specifically involves Usability Experts examining the interface and judging its compliance with recognized usability principles (the "heuristics"). Can be used on paper prototypes and working prototypes (hi and lo fidelity)

## Operation:

- 3 or more usability experts (optimally 4 or more)
  - Best to adhere to some standardized scoring method with allowances for optional input to standardize results (see spreadsheet)
- Applied to the whole page and functional groupings within the page as well as the flow
- The expert goes through the interface several times to ID issues at each level of the interface (Home page, Task orientation, Navigation & IA, Forms & Data Entry, Trust & Credibility, Writing & Content Quality, Page Layout & Visual Design, Search and Help/Feedback/Error)
- Experts 'score' various concepts W/I the check list (if used) to have a concrete measure of good vs bad W/I each topical area.

# Areas Covered (userfocus)

The Home Page; 19 Checkpoints

- The items on the home page are clearly focused on users' key tasks ("featuritis" has been avoided)
- The home page contains a search input box
- **Product categories are provided and clearly visible on the homepage**
- Useful content is presented on the home page or within one click of the home page
- The home page shows good examples of real site content
- Links on the home page begin with the most important keyword (e.g. "Sun holidays" not "Holidays in the sun")
- There is a short list of items recently featured on the homepage, supplemented with a link to archival content
- Navigation areas on the home page are not over-formatted and users will not mistake them for adverts
- The value proposition is clearly stated on the home page (e.g. with a tagline or welcome blurb)
- The home page contains meaningful graphics, not clip art or pictures of models
- Navigation choices are ordered in the most logical or task-oriented manner (with the less important corporate information at the bottom)
- **The title of the home page will provide good visibility in search engines like Google**
- All corporate information is grouped in one distinct area (e.g. "About Us")
- Users will understand the value proposition
- **By just looking at the home page, the first time user will understand where to start**
- The home page shows all the major options

# Areas Covered (userfocus)

## Task Orientation; 43 Checkpoints

- The site is free from irrelevant, unnecessary and distracting information
- Excessive use of scripts, applets, movies, audio files, graphics and images has been avoided
- The site avoids unnecessary registration
- **The critical path (e.g. purchase, subscription) is clear, with no distractions on route**
- Information is presented in a simple, natural and logical order
- The number of screens required per task has been minimised
- The site requires minimal scrolling and clicking
- The site correctly anticipates and prompts for the user's probable next activity
- When graphs are shown, users have access to the actual data (e.g. numeric annotation on bar charts)
- Activities allocated to the user or the computer take full advantage of the strengths of each (look for actions that can be done automatically by the site, e.g. postcode lookup)
- **Users can complete common tasks quickly**
- Items can be compared easily when this is necessary for the task (e.g. product comparisons)
- The task sequence parallels the user's work processes
- The site makes the user's work easier and quicker than without the system
- **The most important and frequently used topics, features and functions are close to the centre of the page, not in the far left or right margins**
- The user does not need to enter the same information more than once
- Important, frequently needed topics and tasks are close to the 'surface' of the web site
- Typing (e.g. during purchase) is kept to an absolute minimum, with accelerators ("one-click") for return users
- The path for any given task is a reasonable length (2-5 clicks)
- When there are multiple steps in a task, the site displays all the steps that need to be completed and provides feedback on the user's current position in the workflow
- The site's privacy policy is easy to find, especially on pages that ask for personal information, and the policy is simple and clear
- **Users of the site do not need to remember information from place to place**
- The use of metaphors is easily understandable by the typical user
- **Data formats follow appropriate cultural conventions (e.g. '-' for phone numbers)**
- Details of the software's internal workings are not exposed to the user
- The site caters for users with little prior experience of the web
- The site makes it easy for users to explore the site and try out different options before committing themselves
- A typical first-time visitor can do the most common tasks without assistance
- When they return to the site, users will remember how to carry out the key tasks
- The functionality of novel device controls is obvious
- On the basket page, there is a highly visible 'Proceed to checkout' button at the top and bottom of the page
- Important calls to action, like 'Add to basket', are highly visible
- **Action buttons (such as "Submit") are always invoked by the user, not automatically invoked by the system when the last field is completed**
- Command and action items are presented as buttons (not, for example, as hypertext links)
- If the user is half-way through a transaction and quits, the user can later return to the site and continue from where he left off
- When a page presents a lot of information, the user can sort and filter the information
- If there is an image on a button or icon, it is relevant to the task
- The site prompts the user before automatically logging off the user, and the time out is appropriate
- Unwanted features (e.g. Flash animations) can be stopped or skipped
- The site is robust and all the key features work (i.e. there are no javascript exceptions, CGI errors or broken links)
- **The site supports novice and expert users by providing different levels of explanation (e.g. in help and error messages)**
- The site allows users to rename objects and actions in the interface (e.g. naming delivery addresses or accounts)
- The site allows the user to customize operational time parameters (e.g. time until automatic logout)

# Areas Covered (userfocus)

Navigation & IA; 28 Checkpoints

- **There is a convenient and obvious way to move between related pages and sections and it is easy to return to the home page**
- The information that users are most likely to need is easy to navigate to from most pages
- Navigation choices are ordered in the most logical or task-oriented manner
- **The navigation system is broad and shallow (many items on a menu) rather than deep (many menu levels)**
- The site structure is simple, with a clear conceptual model and no unnecessary levels
- **The major sections of the site are available from every page (persistent navigation) and there are no dead ends**
- Navigation tabs are located at the top of the page, and look like clickable versions of real-world tabs
- There is a site map that provides an overview of the site's content
- The site map is linked to from every page
- The site map provides a concise overview of the site, not a rehash of the main navigation or a list of every single topic
- Good navigational feedback is provided (e.g. showing where you are in the site)
- Category labels accurately describe the information in the category
- Links and navigation labels contain the "trigger words" that users will look for to achieve their goal
- Terminology and conventions (such as link colours) are (approximately) consistent with general web usage

- **Links look the same in the different sections of the site**
- Product pages contain links to similar and complementary products to support cross-selling
- The terms used for navigation items and hypertext links are unambiguous and jargon-free
- Users can sort and filter catalogue pages (e.g. by listing in price order, or showing 'most popular')
- There is a visible change when the mouse points at something clickable (excluding cursor changes)
- Important content can be accessed from more than one link (different users may require different link labels)
- **Navigation-only pages (such as the home page) can be viewed without scrolling**
- Hypertext links that invoke actions (e.g. downloads, new windows) are clearly distinguished from hypertext links that load another page
- The site allows the user to control the pace and sequence of the interaction
- There are clearly marked exits on every page allowing the user to bale out of the current task without having to go through an extended dialog
- The site does not disable the browser's "Back" button and the "Back" button appears on the browser toolbar on every page
- **Clicking the back button always takes the user back to the page the user came from**
- A link to both the basket and checkout is clearly visible on every page
- If the site spawns new windows, these will not confuse the user (e.g. they are dialog-box sized and can be easily closed)

# Areas Covered (userfocus)

Forms and Data Entry; 22 Checkpoints

- Fields in data entry screens contain default values when appropriate and show the structure of the data and the field length
- When a task involves source documents (such as a paper form), the interface is compatible with the characteristics of the source document
- The site automatically enters field formatting data (e.g. currency symbols, commas for 1000s, trailing or leading spaces). Users do not need to enter characters like £ or %.
- Field labels on forms clearly explain what entries are desired
- **Text boxes on forms are the right length for the expected answer**
- There is a clear distinction between “required” and “optional” fields on forms
- **The same form is used for both logging in and registering (i.e. it's like Amazon)**
- Forms pre-warn the user if external information is needed for completion (e.g. a passport number)
- Questions on forms are grouped logically, and each group has a heading
- Fields on forms contain hints, examples or model answers to demonstrate the expected input
- When field labels on forms take the form of questions, the questions are stated in clear, simple language
- **Pull-down menus, radio buttons and check boxes are used in preference to text entry fields on forms (i.e. text entry fields are not overused)**
- With data entry screens, the cursor is placed where the input is needed
- Data formats are clearly indicated for input (e.g. dates) and output (e.g. units of values).
- Users can complete simple tasks by entering just essential information (with the system supplying the non-essential information by default)
- **Forms allow users to stay with a single interaction method for as long as possible (i.e. users do not need to make numerous shifts from keyboard to mouse to keyboard).**
- The user can change default values in form fields
- Text entry fields indicate the amount and the format of data that needs to be entered
- Forms are validated before the form is submitted
- With data entry screens, the site carries out field-level checking and form-level checking at the appropriate time
- **The site makes it easy to correct errors (e.g. when a form is incomplete, positioning the cursor at the location where correction is required)**
- There is consistency between data entry and data display
- Labels are close to the data entry fields (e.g. labels are right justified)



# Areas Covered (userfocus)

Trust and Credibility; 13 Checkpoints

- The content is up-to-date, authoritative and trustworthy
- The site contains third-party support (e.g. citations, testimonials) to verify the accuracy of information.
- It is clear that there is a real organization behind the site (e.g. there is a physical address or a photo of the office)
- The company comprises acknowledged experts (look for credentials)
- **The site avoids advertisements, especially pop-ups.**
- Delivery costs are highlighted at the very beginning of checkout
- The site avoids marketing waffle
- **Each page is clearly branded so that the user knows he is still in the same site**
- It is easy to contact someone for assistance and a reply is received quickly
- The content is fresh: it is updated frequently and the site includes recent content
- **The site is free of typographic errors and spelling mistakes**
- The visual design complements the brand and any offline marketing messages
- There are real people behind the organization and they are honest and trustworthy (look for bios)

# Areas Covered (userfocus)

Page Layout and Visual Design; 38 Checkpoints

- The layout helps focus attention on what to do next
- **On all pages, the most important information (such as frequently used topics, features and functions) is presented on the first screenful of information (“above the fold”)**
- **The site can be used without scrolling horizontally**
- Things that are clickable (like buttons) are obviously pressable
- Items that aren't clickable do not have characteristics that suggest that they are
- The functionality of buttons and controls is obvious from their labels or from their design
- Clickable images include redundant text labels (i.e. there is no 'mystery meat' navigation)
- Hypertext links are easy to identify without needing to 'minesweep' (e.g. underlined)
- Fonts are used consistently
- The relationship between controls and their actions is obvious
- Icons and graphics are standard and/or intuitive (concrete and familiar)
- There is a clear visual "starting point" to every page
- Each page on the site shares a consistent layout
- Pages on the site are formatted for printing, or there is a printer-friendly version
- Buttons and links show that they have been clicked
- GUI components (like radio buttons and check boxes) are used appropriately
- Fonts are readable
- The site avoids italicized text and uses underlining only for hypertext links
- There is a good balance between information density and use of white space
- The site is pleasant to look at
- Icons are visually and conceptually distinct yet still harmonious (clearly part of the same family)
- Related information and functions are clustered together, and each group can be scanned in a single fixation (5-deg, about 4.4cm diam circle on screen)
- Pages are free of "scroll stoppers" (headings or page elements that create the illusion that users have reached the top or bottom of a page when they have not)
- The site avoids extensive use of upper case text
- The site has a consistent, clearly recognisable look and feel that will engage users
- Saturated blue is avoided for fine detail (e.g. text, thin lines and symbols)
- Colour is used to structure and group items on the page
- Graphics will not be confused with banner ads
- Emboldening is used to emphasise important topic categories
- **On content pages, line lengths are neither too short (<50 characters per line) nor too long (>100 characters per line) when viewed in a standard browser width window**
- Pages have been designed to an underlying grid, with items and widgets aligned both horizontally and vertically
- Meaningful labels, effective background colors and appropriate use of borders and white space help users identify a set of items as a discrete functional block
- The colors work well together and complicated backgrounds are avoided
- Individual pages are free of clutter and irrelevant information
- **Standard elements (such as page titles, site navigation, page navigation, privacy policy etc.) are easy to locate**
- The organisation's logo is placed in the same location on every page, and clicking the logo returns the user to the most logical page (e.g. the home page)
- Attention-attracting features (such as animation, bold colours and size differentials) are used sparingly and only where relevant

# Areas Covered (userfocus)

Search; 20 Checkpoints

- The default search is intuitive to configure (no Boolean operators)
- The search results page shows the user what was searched for and it is easy to edit and resubmit the search
- Search results are clear, useful and ranked by relevance
- The search results page makes it clear how many results were retrieved, and the number of results per page can be configured by the user
- If no results are returned, the system offers ideas or options for improving the query based on identifiable problems with the user's input
- **The search engine handles empty queries gracefully**
- The most common queries (as reflected in the site log) produce useful results
- The search engine includes templates, examples or hints on how to use it effectively
- The site includes a more powerful search interface available to help users refine their searches (preferably named "revise search" or "refine search", not "advanced search")
- The search results page does not show duplicate results (either perceived duplicates or actual duplicates)
- The search box is long enough to handle common query lengths
- Searches cover the entire web site, not a portion of it
- If the site allows users to set up a complex search, these searches can be saved and executed on a regular basis (so users can keep up-to-date with dynamic content)
- **The search interface is located where users will expect to find it (top right of page)**
- **The search box and its controls are clearly labelled (multiple search boxes can be confusing)**
- The site supports people who want to browse and people who want to search
- The scope of the search is made explicit on the search results page and users can restrict the scope (if relevant to the task)
- The search results page displays useful meta-information, such as the size of the document, the date that the document was created and the file type (Word, pdf etc.)

# Areas Covered (userfocus)

Help Feedback & Error Tolerance; 37 Checkpoints

- The FAQ or on-line help provides step-by-step instructions to help users carry out the most important tasks
- It is easy to get help in the right form and at the right time
- Prompts are brief and unambiguous
- The user does not need to consult user manuals or other external information to use the site
- The site uses a customised 404 page, which includes tips on how to find the missing page and links to “Home” and Search
- **The site provides good feedback (e.g. progress indicators or messages) when needed (e.g. during checkout)**
- **Users are given help in choosing products**
- User confirmation is required before carrying out potentially “dangerous” actions (e.g. deleting something)
- Confirmation pages are clear
- **Error messages contain clear instructions on what to do next**
- Immediately prior to committing to the purchase, the site shows the user a clear summary page and this will not be confused with a purchase confirmation page
- When the user needs to choose between different options (such as in a dialog box), the options are obvious
- The site keeps users informed about unavoidable delays in the site’s response time (e.g. when authorising a credit card transaction)
- Error messages are written in a non-derisory tone and do not blame the user for the error
- Pages load quickly (5 seconds or less)
- **The site provides immediate feedback on user input or actions**
- The user is warned about large, slow-loading pages (e.g. “Please wait...”), and the most important information appears first
- Where tooltips are used, they provide useful additional help and do not simply duplicate text in the icon, link or field label
- When giving instructions, pages tell users what to do rather than what to avoid doing
- The site shows users how to do common tasks where appropriate (e.g. with demonstrations of the site's functionality)

# Areas Covered (userfocus)

Help Feedback & Error Tolerance; 37 Checkpoints

- The site provides feedback (e.g. “Did you know?”) that helps the user learn how to use the site
- The site provides context sensitive help
- Help is clear and direct and simply expressed in plain English, free from jargon and buzzwords
- The site provides clear feedback when a task has been completed successfully
- Important instructions remain on the screen while needed, and there are no hasty time outs requiring the user to write down information
- **Fitts' Law is followed (the distance between controls and the size of the controls is appropriate, with size proportional to distance)**
- There is sufficient space between targets to prevent the user from hitting multiple or incorrect targets
- **There is a line space of at least 2 pixels between clickable items**
- The site makes it obvious when and where an error has occurred (e.g. when a form is incomplete, highlighting the missing fields)
- The site uses appropriate selection methods (e.g. pull-down menus) as an alternative to typing
- **The site does a good job of preventing the user from making errors**
- The site prompts the user before correcting erroneous input (e.g. Google's “Did you mean...”)
- The site ensures that work is not lost (either by the user or site error)
- Error messages are written in plain language with sufficient explanation of the problem
- When relevant, the user can defer fixing errors until later in the task
- The site can provide more detail about error messages if required
- **It is easy to “undo” (or “cancel”) and “redo” actions**

# User Focus Expert Evaluation Results

Output:

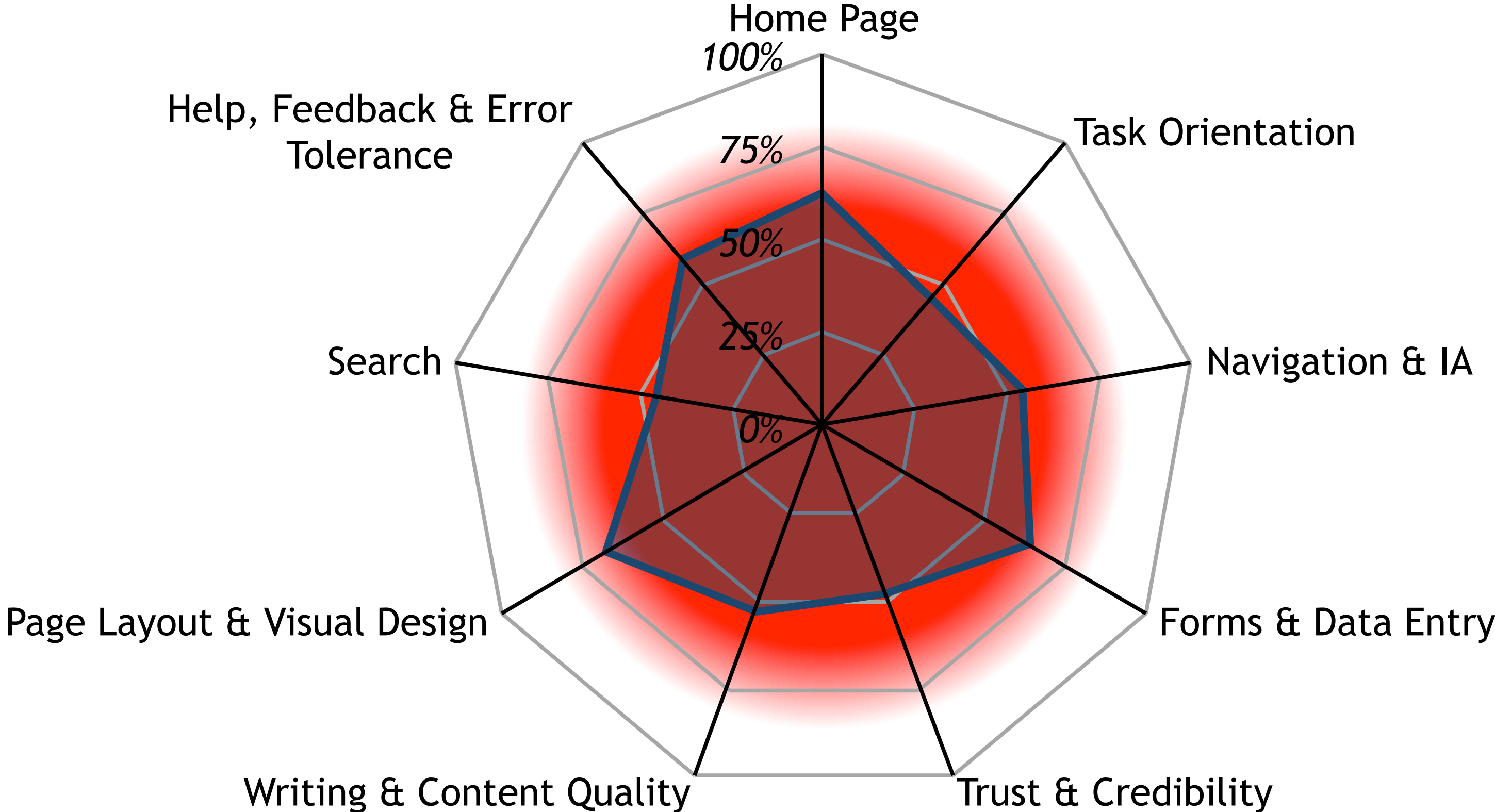
Summary of results				
	Raw score	# Questions	# Answers	Score
Home Page	0	20	0	
Task Orientation	0	44	0	
Navigation & IA	0	29	0	
Forms & Data Entry	0	23	0	
Trust & Credibility	0	13	0	
Writing & Content Quality	0	23	0	
Page Layout & Visual Design	0	38	0	
Search	0	20	0	
Help, Feedback & Error Tolerance	0	37	0	
Overall score		247	0	

- Hint: The 247 Checkpoints ARE the points that should be paid attention to when designing an interface
- ATT usability uses the USERFOCUS Checklist to evaluate Tek Designs



# And your results could wow audiences...

Output:



# HOMWORK

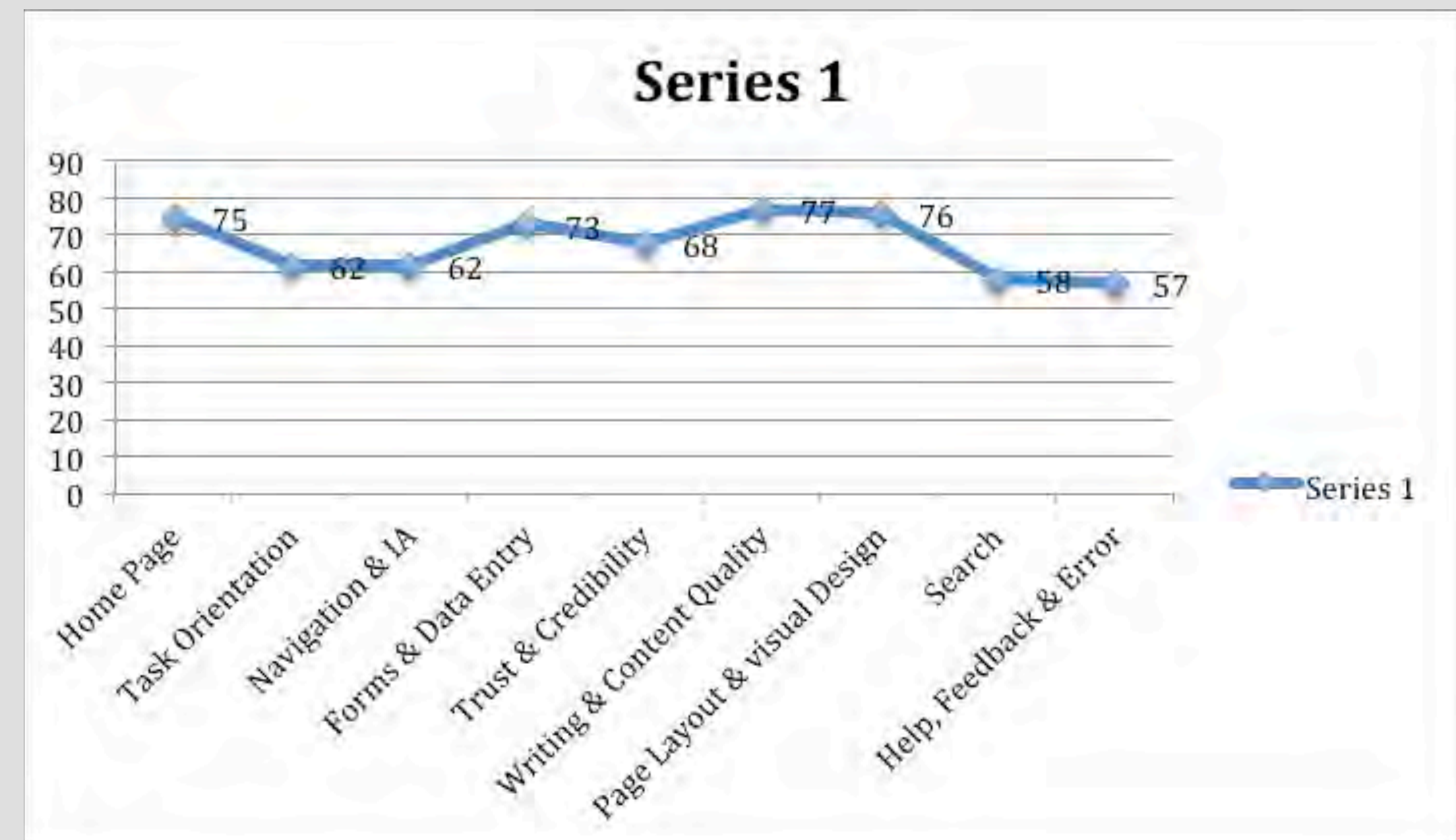
Download the UserFocus Heuristic Spreadsheet and review it

Run YOUR OWN PERSONAL Heuristic on the [tekzenit.com](http://tekzenit.com) website tonight w/ comments

In the morning combine your scores to get an average of all your assessments and Plot them per person on one graph so that each person is represented by a line. We will start back at that point after you've completed that assignment (9:45am)

<b>Heuristic Analysis of:</b>	<b>1. Score %</b>	<b>2. Score %</b>	<b>3. Score%</b>	<b>Average Score %</b>
Home Page	72%	65%	88%	75
Task Orientation	68%	47%	72%	62
Navigation & IA	66%	45%	76%	62
Forms & Data Entry	82%	70%	68%	73
Trust & Credibility	85%	59%	60%	68
Writing & Content Quality	83%	66%	83%	77
Page Layout & Visual Design	75%	65%	88%	76
Search	45%	53%	75%	58
Help, Feedback & Error Tolerance	62%	44%	64%	57
Overall score	71%	57%	75%	68

You will present 3 things; spider, plot, and spreadsheet like to the left & below.





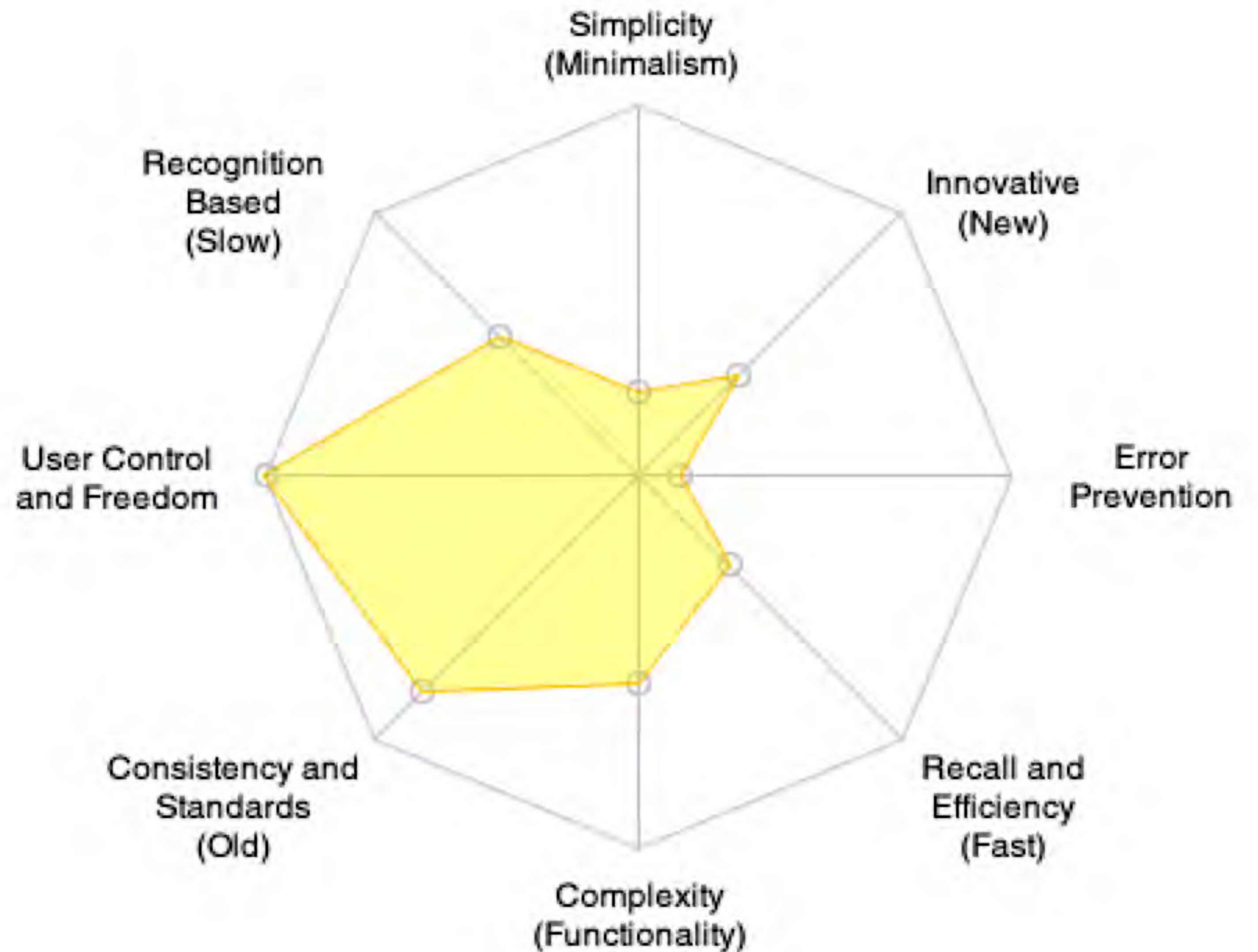
# HOMework

## Class Presentation

# Followup

## Be aware.....

- Context is important
- Design Direction is important
- Some Heuristics, on the surface, seem to be opposed
- Is your application 'mission-critical'?
- Is it for leisure or work?
- How complex is the job that the user has to do?



# Early Prototype Testing

(Low Fidelity)

## Agenda

Paper Prototyping

Tree Test (covered previously)

Expectancy Test

Performance Test

Visual Affordance Test

**August 2014**

# Paper Prototyping

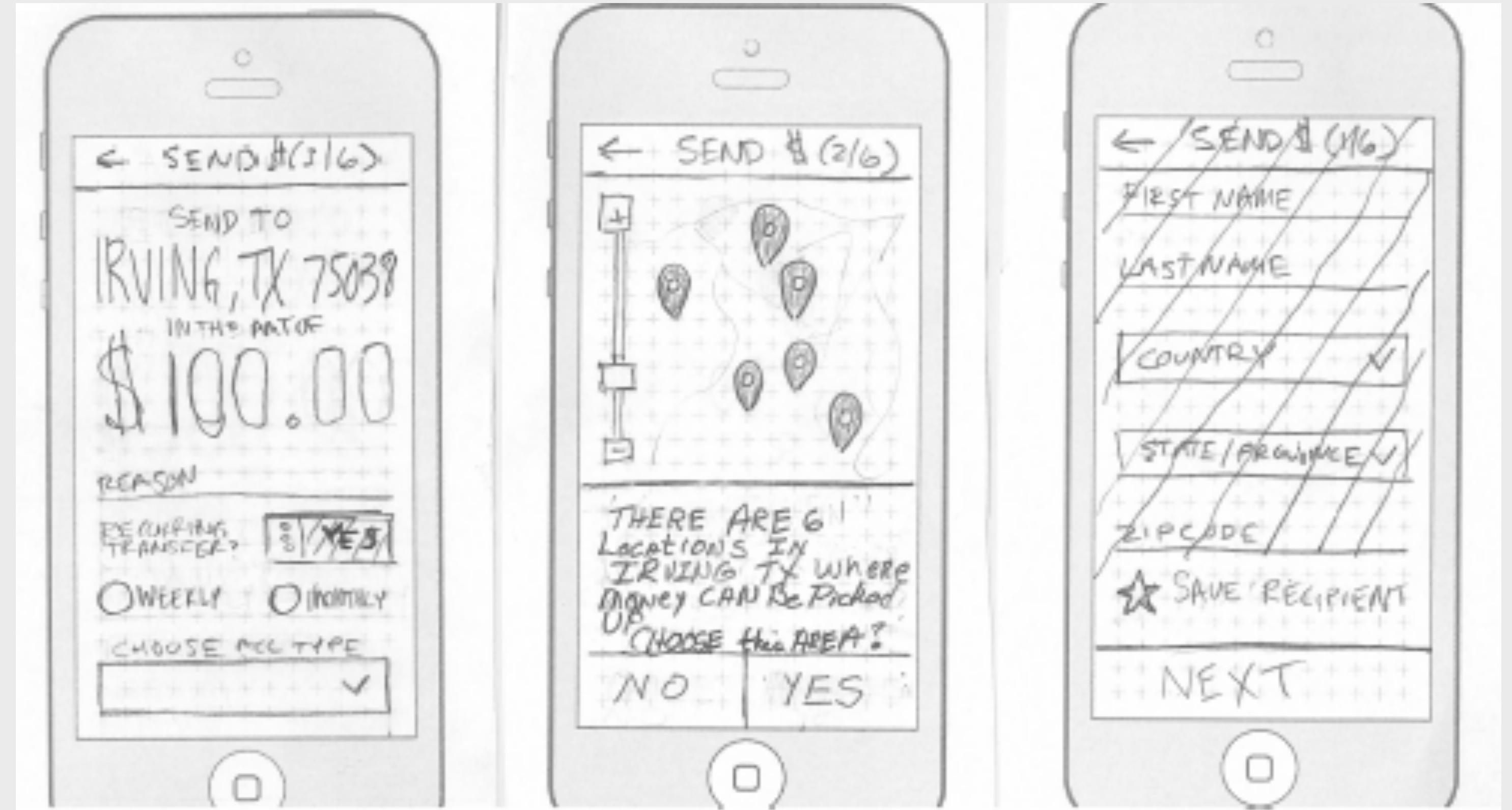
Very Early in the process

August 2014

## *Early Prototype Testing - Paper Prototyping*

- Uses paper prototypes & low fidelity (wireframe) prototypes
- Evaluates
  - Usability & efficiency of high-level navigational design
  - 'Fit' of the prototype to the users' mental model
  - Primary content and organization
- Builds a strong foundation for the detailed design
- Is the most cost-effective usability test
- Helps evaluate visual presentation and branding

# Prior Project ...



# Output ...

## Survey Data (post task)

	User 1	User 2	User 3	User 4	User 5	Average	Notes
<i>How attractive do you think the app is?(1=ugly 7=very attractive)</i>	5	7	6	4	6	5.6	Value indicates it could be more attractive
<i>How difficult is it to use the app? (1=easy 7=hard)</i>	1	6	4	1	6	3.6	Value indicates it could be easier, is just 'average'
<i>Have you created a permanent account? (validation of understanding what they did)</i>	F	T	T	T	T	N/A	20% of users were confused as to whether they created a perm account
<i>None of the personal info you've entered has been saved? (validation of understanding what they did)</i>	F	F	F	F	F	N/A	Users understood they had saved personal info on MoneyGram 100%
<i>How secure do you feel when using the app? (1=not secure 7=very secure)</i>	4	6	7	5	7	5.8	
<i>Why not perceived to be secure?</i>	Don't see reassurances			App doesn't look trustworthy		N/A	Items impacting perceived trustworthiness include reassurances such as SSL, Verified, encryption and visual professionalism
<i>How frequently would you use this app? (1=would not use 7=use all the time)</i>	4	7	1	4	1	3.4	Tendency to NOT USE this app (below average use score)
<i>Why not use?</i>	Use my bank only		Don't trust money sending apps	Would use something else	Don't know the Brand		Items impacting decision not to use include Would use my bank only, don't trust money sending apps, Don't know the MoneyGram Brand and would use something else

## *Output ...*

- Survey data
- User observations
- Designer understanding of user issues
- Design changes
- Mock ups for version 2 of the prototype



# ***Benefits***

- Fast Development Time
- Ease of change
- Designer involvement
- Focus on basic design
- No technology constraints
- Users more willing to critique something in progress
- Designers more willing to take criticism

# Expectancy Testing

When there are visual comps ready


August 2014

## *Expectancy Test*

- Asks for participants' initial understanding of the purpose, value, benefit of the application or web site
- Determines whether the design elements communicate effectively
- Usually done prior to a performance test procedure
- Asks questions such as:
  - *“Just look at this interface and w/o using the mouse or keyboard, tell me what you think it does.”*
  - *“What would you expect to find under each of the major categories?”*

# Expectancy Test

“What would you expect to find under each of the major categories?”


Continue to the new My Verizon >>

- 1
- 2
- 3
- 4

## Simple. Fast. Most importantly yours.

The new My Verizon is Here!

Now it's easier than ever to manage every aspect of your account. Because now you're never more than a few clicks away from the things you need to do, and the things you want to do. Scroll down to explore.

SCROLL DOWN 

# Expectancy Test

“What would you expect to find under each of the major categories?”



[Continue to the new My Verizon >>](#)

## Simple. Fast.

1 Most importantly yours.

2 Your account

3 Manage your services

4 Pay your Bill

Find answers

the every how you're from the things you need to do, and the things you want to do. Scroll down to explore.

SCROLL DOWN 



# Performance Testing

HTML click testing\*

August 2014

# Performance Test

- Evaluate the navigation design
- Use test scenarios to watch, observe, and listen to user as they interact
- Focus on Performance, not opinions

*Find out if your information is secure  
Look for a job  
Create a profile*

The screenshot displays the iinVIEW website homepage. At the top, there is a navigation bar with the iinVIEW logo and links for 'About Us', 'Newsroom', 'Blog', and 'Contact'. On the right side of the navigation bar are 'LOG IN' and 'SIGN UP NOW' buttons. Below the navigation bar is a search section titled 'WHAT ARE YOU LOOKING FOR?' with a 'CAREERS' dropdown menu, a search input field containing 'Search iinView Jobs', a 'Location' dropdown, a '20 miles' radius selector, and a 'SEARCH' button. The main content area features two promotional banners. The left banner is titled 'ARE YOU A JOB SEEKER?' with the subtext 'Find the Perfect Opportunity.' and a 'LEARN MORE' button. The right banner is titled 'LOOKING FOR EMPLOYEES?' with the subtext 'Get Introduced to the Top Candidates.' and a 'LEARN MORE' button. Below these banners is a large orange section titled 'THE KEYS TO BEING IINVIEW' containing three icons and their corresponding text: an eye icon for 'INCREASED PROFESSIONAL VISIBILITY', a video player icon for 'EASY TO USE VIDEO INTERVIEWS', and a padlock icon for 'SECURE PROFESSIONAL NETWORK'. The footer is dark grey and contains links for 'Home', 'About Us' (with sub-links 'History' and 'Executive Team'), 'Newsroom' (with sub-links 'Press Releases', 'News Articles', and 'Media Kit'), 'Blog' (with sub-links 'Field Progression' and 'From Temp to Full Hiring?'), and 'Contact' (with sub-link 'Privacy'). Social media icons for Facebook, Twitter, and YouTube are on the left, and the iinVIEW logo is on the right. A copyright notice '© COPYRIGHT IINVIEW 2012 ALL RIGHTS RESERVED - PATENT PENDING' is centered in the footer.

# Visual Affordance Test

When there are visual comps ready

August 2014



# Visual Affordance Testing

## What is clickable and what is not?

The website otherwise not only seems to be completely Flash-based, but also features a huge intro video (without any content), Mystery Meat Navigation (animated with flying items, no less) and, according to my tests, it does not work in Firefox, does not work in Opera and does not work in Safari.



# Brand Definition Test

When there are visual comps ready

August 2014

# Brand Definition Test

How do People react to a design?



Intriguing	16%	Disinteresting	2%
Exciting	2%	Boring	2%
Sophisticated		Common	
Quality		Cheap	2%
Strong	4%	Weak	
Youthful	7%	Old Fashioned	
Comfortable		Edgy	12%
Trustworthy		Unreliable	
Energetic	9%	Lethargic	
Unique	16%	Common	
Attractive	7%	Ugly	4%
Modern	12%	Traditional	2%

This was done in house using the paired comparison technique across several designs

# Brand Definition Test

How do People react to a design?



Intriguing	12%	Disinteresting	4%
Exciting	2%	Boring	4%
Sophisticated	10%		
Quality	6%	Cheap	
Strong	8%	Weak	4%
Youthful	6%	Old Fashioned	
Comfortable	2%	Edgy	2%
Trustworthy	4%	Unreliable	
Energetic	6%	Lethargic	
Unique	6%	Common	
Attractive	6%	Ugly	
Modern	13%	Traditional	6%

This was done in house using the paired comparison technique across several designs

# Brand Definition Test

How do People react to a design?



Intriguing		Disinteresting	9%
Exciting		Boring	9%
Sophisticated	7%		
Quality	2%	Cheap	2%
Strong	7%	Weak	2%
Youthful		Old Fashioned	7%
Comfortable	7%	Edgy	2%
Trustworthy		Unreliable	
Energetic	2%	Lethargic	
Unique		Common	18%
Attractive	2%	Ugly	2%
Modern	11%	Traditional	9%



This was done in house using the paired comparison technique across several designs

# Brand Definition Test

How would YOU react to this design?

Text Only Version Tuesday, December 12, 2006 | 7:12 PM

**THE UNIVERSITY OF TEXAS AT EL PASO** Main Navigation Department/Program Search GO

Home

**President's Welcome**  
 About UTEP and Region  
 Academic Programs  
 Admissions  
 Alumni  
 Athletics  
 Arts and Culture  
 Jobs  
 Graduate Studies  
 Library  
 Research  
 Student Life  
 Directory and Maps

**Giving to UTEP**

Show me links for: UTEP Web Search GO

**University News** For the media

Alumni Ho

View Webmail, WebCT and Goldmine at My.UTEP.edu!

**Calendar of Events**  
 Tuesday, December 12, 2006

- Blanket Drive
- Luis Jiménez Tributes
- Wintermester Tutoring Schedule
- FEMTEC 2006

**More of Today's Events**  
 Upcoming Events - December 2006

- Alumni Holiday Social
- "A Christmas Carol"
- Season of Lights
- American Chemical Society D...
- "The Nutcracker"

**More Upcoming Events**

**Highlights**

PAYING for COLLEGE NOVA FALL 2006 ONLINE MINER e-NOTES

State Reports | Customer Service Statement | State of Texas | Statewide Search | Privacy Policy | Open Records | Site Feedback

500 West University Avenue | El Paso, Texas 79968 | (915) 747-5000

## *Summary ....*

- Early Prototypes test navigation and content organization in the most cost effective way
  - There are five types of tests in the early prototype stages
    - Expectancy
    - Tree Test
    - Performance
    - Visual Affordance
    - Branding

# Advanced Prototype Testing

For a functional prototype  
(High Fidelity)

## Agenda

Free Exploration  
Eye Tracking  
Performance Test

**August 2014**



Early prototypes are ideal for testing navigation and labeling systems. They focus on structure and organization in the content of a few key pages



Advanced prototypes are ideal for testing early detailed design. They focus on details and interactivity

TERMS & CONDITIONS ▾

## Personal Information ▾

---

### Employment Information ▾

---

Employment Status  
Please Select ▾

Employer Name  
Enter Name

Occupation  
Enter Occupation

Annual Income Why is this needed?  
\$15,000 - \$25,000 ▾

NEXT STEP

---

### Authorized User ▾

---

**Need Help?**

If you're experiencing technical difficulties while applying on our website, please call Technical Support at [1 \(877\) 277-0946](tel:18772770946) 6:00am to 5:00pm PT Monday-Friday

TERMS & CONDITIONS ▾

## Account Verification ▾

---

First Name  
John Doe

Last Name  
John Doe

Costco Membership ID  
\*\*\*\*\*

+ Add an authorized user

NEXT STEP

---

**Need Help?**

If you're experiencing technical difficulties while applying on our website, please call Technical Support at [1 \(877\) 277-0946](tel:18772770946) 6:00am to 5:00pm PT Monday-Friday

## Advanced Prototype Testing (High Fidelity) uses a functioning prototype

- Validates the mental model and navigation
- Evaluates page flow and interaction
- Evaluates detailed design issues
- Uncovers technology issues
- Tests new pages that were not included in the early prototypes
- Tests page-level details including:
  - Content
  - Page Flow
  - Editorial style
  - Page-level controls and interaction
  - Graphics
  - Consistency issues

# Advanced Prototype Testing

## Advantages

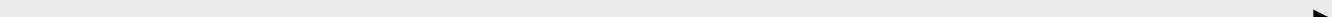
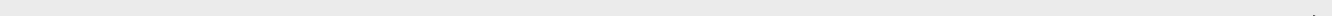
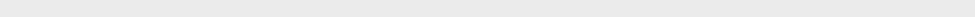
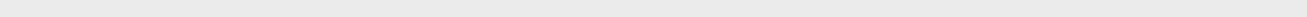
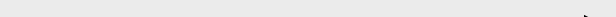
- Greater depth of design feedback
- Presentation close to the real thing
- Allows for greater interaction

## Limitations

- Large development investment
- Ability to easily make changes is limited
- Users less likely to be open and critical

# Advanced Prototype Testing

## To Evaluate

- Actual Usage 
- Actual Usage 
- Organization of Content 
- Navigation Design 
- Layout, Task Flow, Controls and Content 

## Use This

- Free Exploration Test
- Eye Tracking
- Performance Test
- Performance Test
- Performance Test

# Free Exploration Testing

August 2014

# Free Exploration Test

## Participants use the site without scenarios

- More realistic
  - Typical site/software usage
  - Exploration of more areas in a realistic manner
  - Intrinsic (internal) rather than extrinsic (external) motivation
- Allows the facilitator to probe when choices are made
- Can be used prior to test scenarios (but beware of learning effects)

## Do NOT use when

- If navigation is a major focus of the upcoming scenario tests

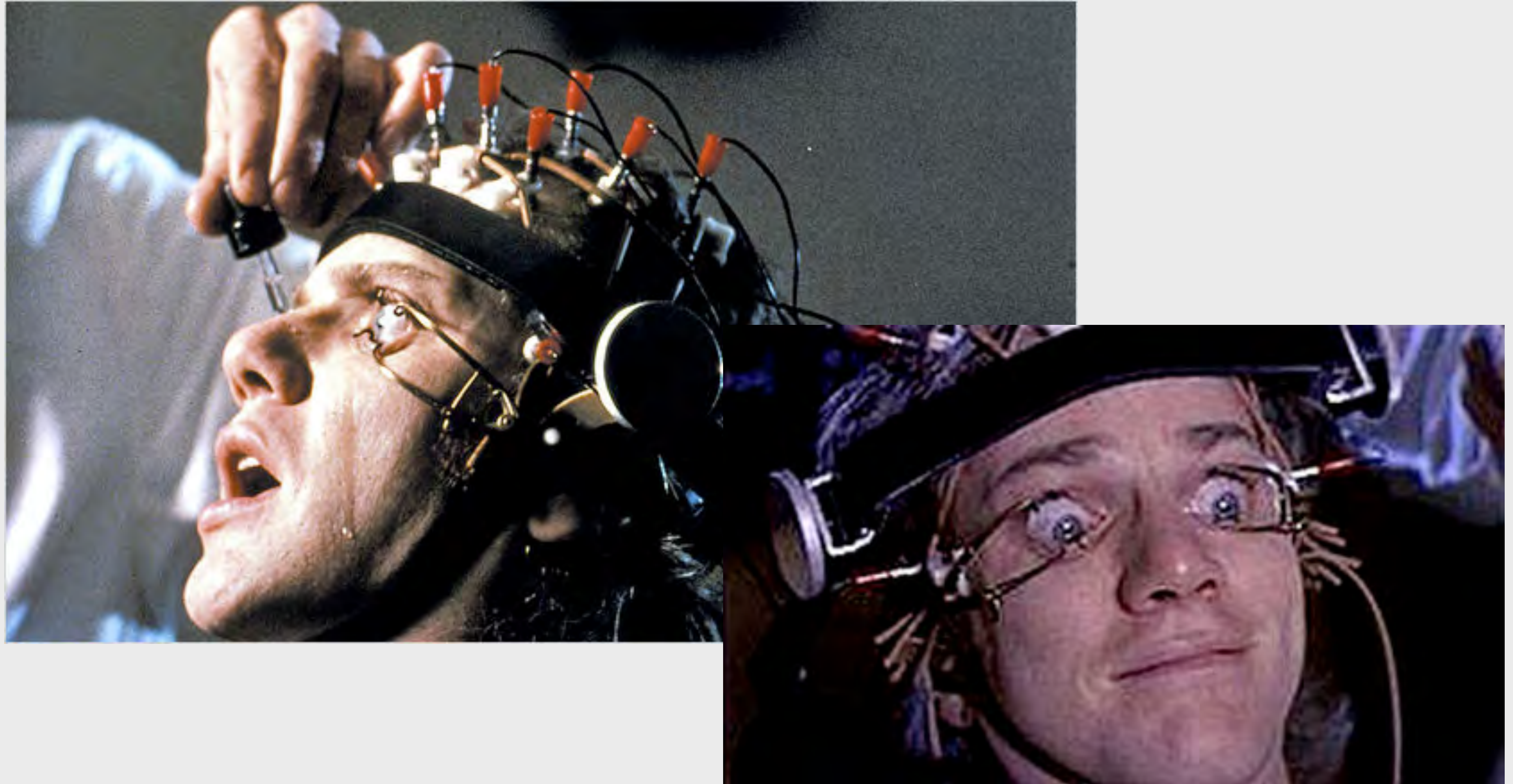
# Eye-Tracking

August 2014



# Eye-Tracking

**It's not just for creepy mad scientists types anymore.....**



# Eye-Tracking

## Remarkably Useful

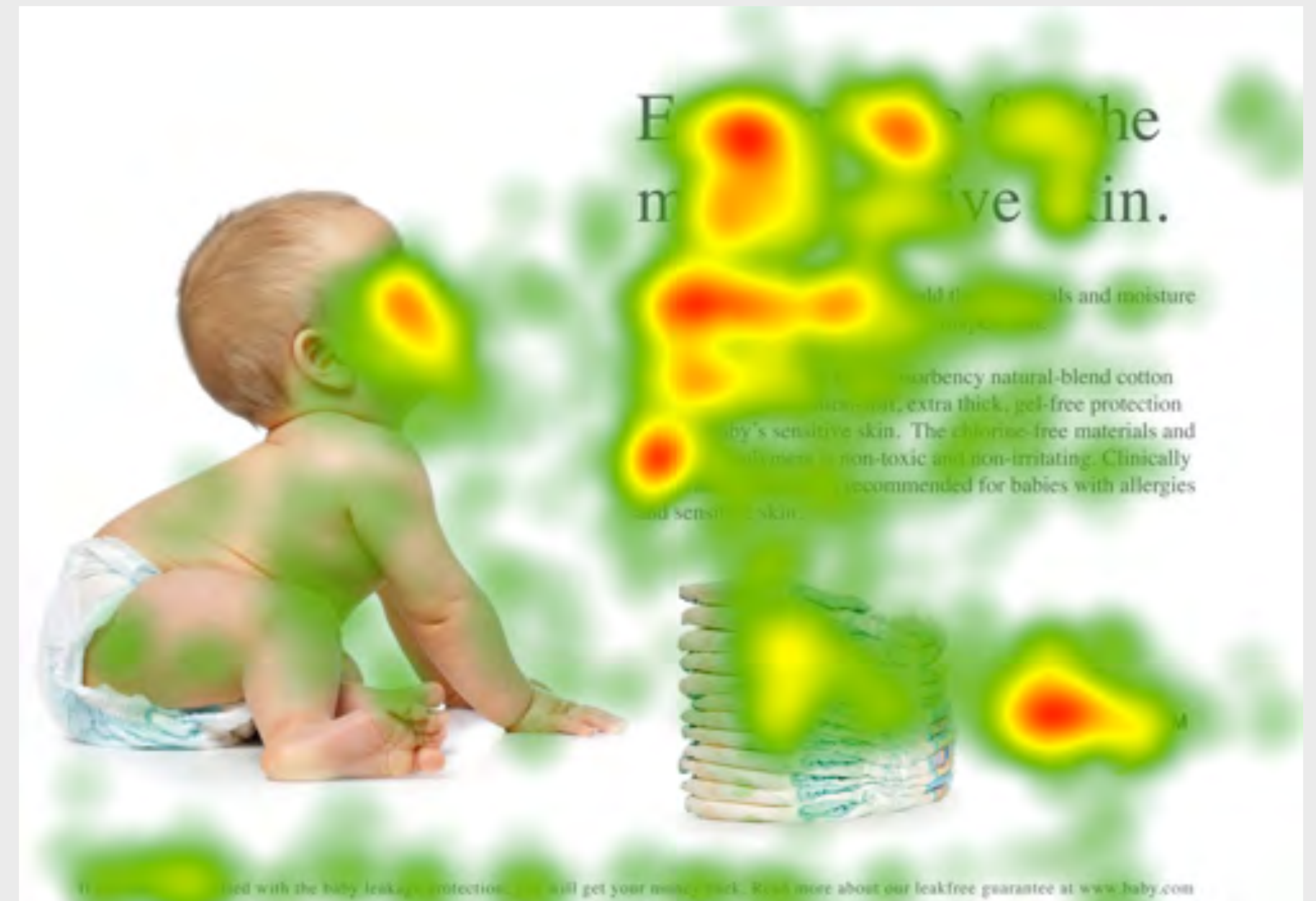
- Total amount of TIME a user looks at an area of interest; Link, Image, Widget (Dwell Time)
  - Level of interest in an area of the screen
- How MANY TIMES did the user pause their eye movement over a specific area of the visual field (Number of fixations)
  - Level of interest
- On AVERAGE, how LONG was the user's pause over an area of interest (Fixation Duration)
- In what ORDER did the user look at things on the screen? (Sequence)
- How LONG did it take the user to NOTICE a particular element on the screen? (Time to First fixation)
- How MANY TIMES did the user REVISIT that particular element on the screen? (Revists)
- What was the percentage of users who had at least one fixation within that area of interest? (Hit Ratio)

Discuss how these would be helpful....

# Eye-Tracking

## What do people look at?

### Static Heatmap



# Why the Difference?

# Eye-Tracking

## What do people look at?

Static Heatmap

# Why the Difference?



# Eye-Tracking

## What do people look at?

### Static Heatmap



Why?

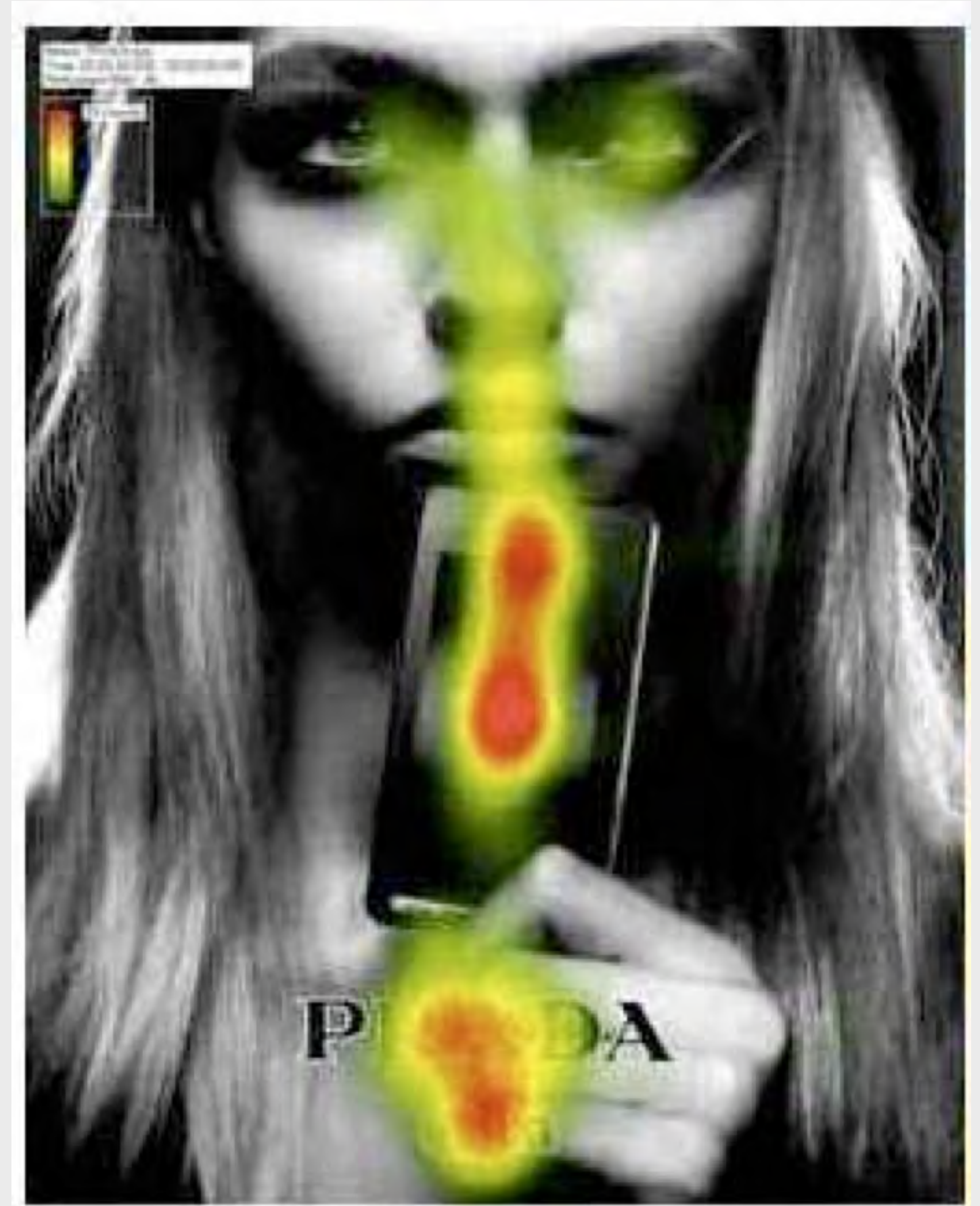


# Eye-Tracking

What do people look at?

Static Heatmap

What's going on here?



# Eye-Tracking

## Example Gaze Plot

**GazePlot**  
 Media: http://zu5eoa.axshare.com/homepage.html (CRC)  
 Time: 00:00:00.000 - 00:00:39.706  
 Participant filter: All Participants  
 Number of participants included: 1/1 (100%)

**iinVIEW** About Us Newsroom Blog Contact LOG IN SIGN UP NOW

**ARE YOU A  
JOB SEEKER?**  
 Find the Perfect Opportunity.

[LEARN MORE](#)

**LOOKING FOR  
EMPLOYEES?**  
 Get Introduced to the Top Candidates.

[LEARN MORE](#)

WHAT ARE YOU LOOKING FOR?

CAREERS  Location  [SEARCH](#)

THE KEYS TO BEING IINVIEW

INCREASED PROFESSIONAL  
VISIBILITY

EASY TO USE VIDEO  
INTERVIEWS

SECURE PROFESSIONAL  
NETWORK

Home About Us History Executive Team Newsroom Press Releases News Articles Media Kit Blog Field Progression From Temp to Full Hiring? Contact Privacy

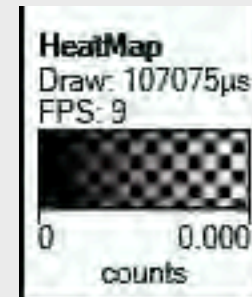
f t v + © COPYRIGHT IINVIEW 2013 ALL RIGHTS RESERVED. IINVIEW IINVIEW **iinVIEW**

# Eye-Tracking

## Example

Revealmap -

Heatmap





# Performance Testing

August 2014

# Performance Test

- Evaluate the design in relation to TASKS
- Use test scenarios to watch, observe, and listen to user as they interact and RECORD METRICS
- Focus on Performance and pathways, not opinions

The screenshot displays the iinVIEW website interface. At the top, there is a navigation bar with the iinVIEW logo and links for About Us, Newsroom, Blog, and Contact. On the right side of the navigation bar, there are buttons for LOG IN and SIGN UP NOW. Below the navigation bar, a search section titled "WHAT ARE YOU LOOKING FOR?" features a "CAREERS" dropdown menu, a search input field with the placeholder "Search iinView Jobs", a "Location" field, a "20 miles" radius selector, and a "SEARCH" button. The main content area is divided into two columns. The left column is titled "ARE YOU A JOB SEEKER?" and includes the subtext "Find the Perfect Opportunity." and a "LEARN MORE" button. The right column is titled "LOOKING FOR EMPLOYEES?" and includes the subtext "Get Introduced to the Top Candidates." and a "LEARN MORE" button. Below these columns, a large orange banner contains the text "THE KEYS TO BEING IINVIEW" and three icons: an eye for "INCREASED PROFESSIONAL VISIBILITY", a play button for "EASY TO USE VIDEO INTERVIEWS", and a padlock for "SECURE PROFESSIONAL NETWORK". The footer is dark gray and contains links for Home, About Us (with sub-links for History and Executive Team), Newsroom (with sub-links for Press Releases, News Articles, and Media Kit), Blog (with sub-links for Field Progression and From Temp to Full Hiring?), and Contact (with a sub-link for Privacy). Social media icons for Facebook, Twitter, and YouTube are also present. A copyright notice at the bottom center reads "© COPYRIGHT IINVIEW 2012 ALL RIGHTS RESERVED - PATENT PENDING". The iinVIEW logo is located in the bottom right corner of the footer.

# Review

August 2014

## What is Wrong here?

### Objective:

Identify possible user friction points in the interactions, tasks, and flows within the [ ] User Experience.

### Methodology:

The team performed a Summative Usability Test with 9 users who provided feedback on the overall usability of key tasks that they performed during 60 minute, in-person, 1:1 sessions in [ ] -

### Timeline:

July 1st: Design/IA Team review end to end flows with UX Research ( [ ] )

July 7th: Detailed comps/wireframes and end to end flows provided by IAs to [ ] to create prototype

July 14th: Prototype review amongst the broader team

July 23rd – 25th: Usability Study conducted in [ ]

August 11th: Detailed Report and Results Presentation

## What is Wrong here?

## Participants

Participant	# Employees	Employee or Owner	Name of Company	Mobile Service Provider	Nature of Business	Age	Gender
1	20-100	Owner	Legacy Ballet	Verizon	Parking Management	26-35	Male
2	10-19	Owner	Window Man Window Cleaning	Metro PCS	Window Cleaning Services	26-35	Female
3	2-9	Owner	Classic Dry Cleaners	AT&T	Dry Cleaning	Under 25	Male
4	20-100	Owner	Total Air and Heat Company	Sprint, AT&T	HVAC	50+	Male
5	10-19	Employee	Legacy National Signs	Verizon	Manufacturing	36-50	Male
6	2-9	Owner	World Trade Consortium	AT&T	Sell Commodities (Retail)	50+	Male
7	10-19	Employee	Raymond Constructions	AT&T	Construction	36-50	Female
8	20-100	Employee	William Stukay	Verizon	Accounting and Finance	36-50	Female
9	20-100	Owner	Cross Roads Diner	AT&T	Restaurant	36-50	Male

One additional participant was recruited. However, she did not complete her entire session and is not included in the results due to being an outlier.

"You need to test with **5 users** to detect approximately **85% of the problems** in an interface, given that the probability a user would encounter the problem is about 31%." [\(Nielsen and Landauer, 1993\)](#)

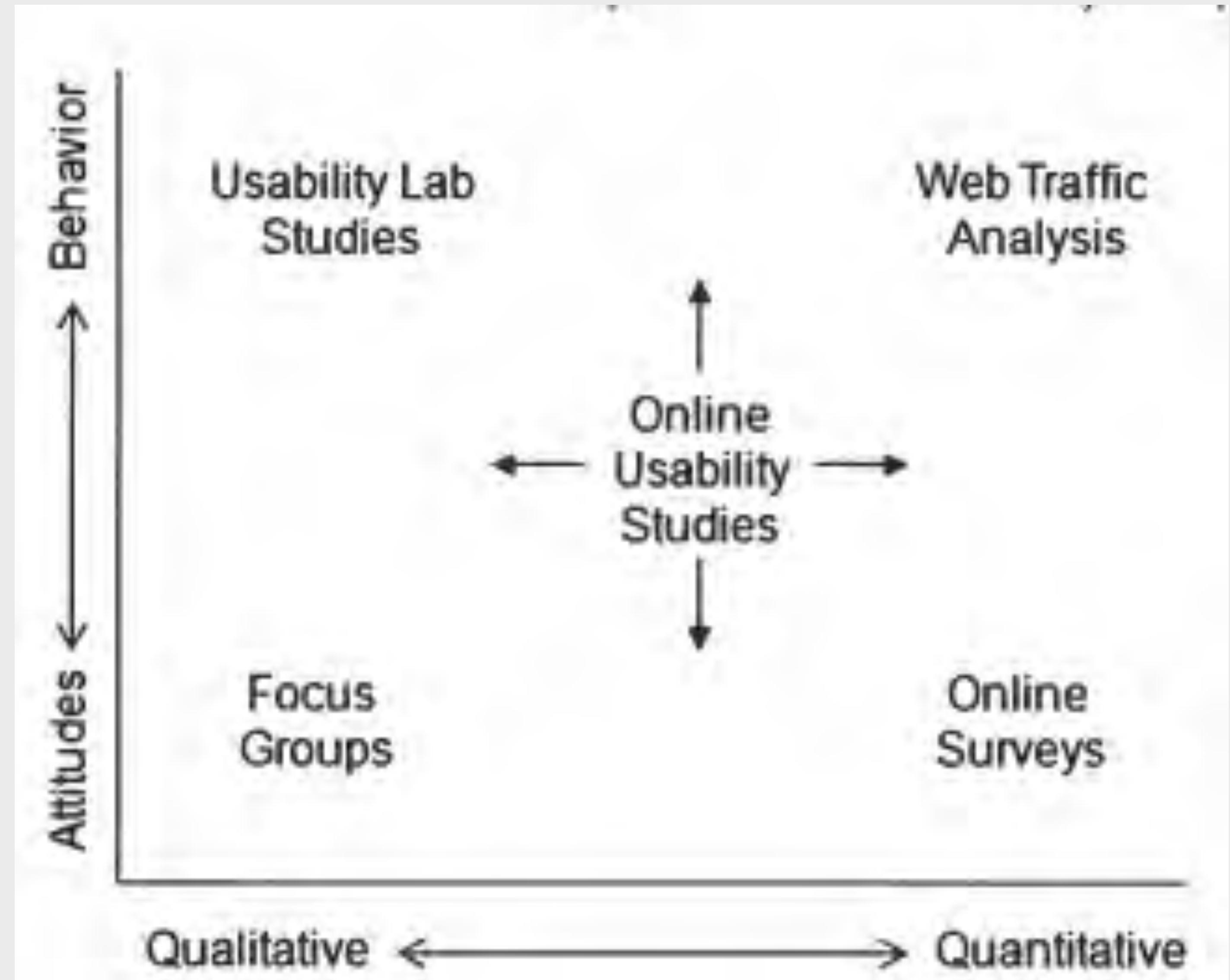
# Online (Remote) vs Lab Testing

August 2014

## Online vs Lab

### More and more can be done online as technology progresses

- Clickmaps / Heatmaps
- Surveys
- Some low end eye tracking
- Performance Data
- ...and more



# Online vs Lab

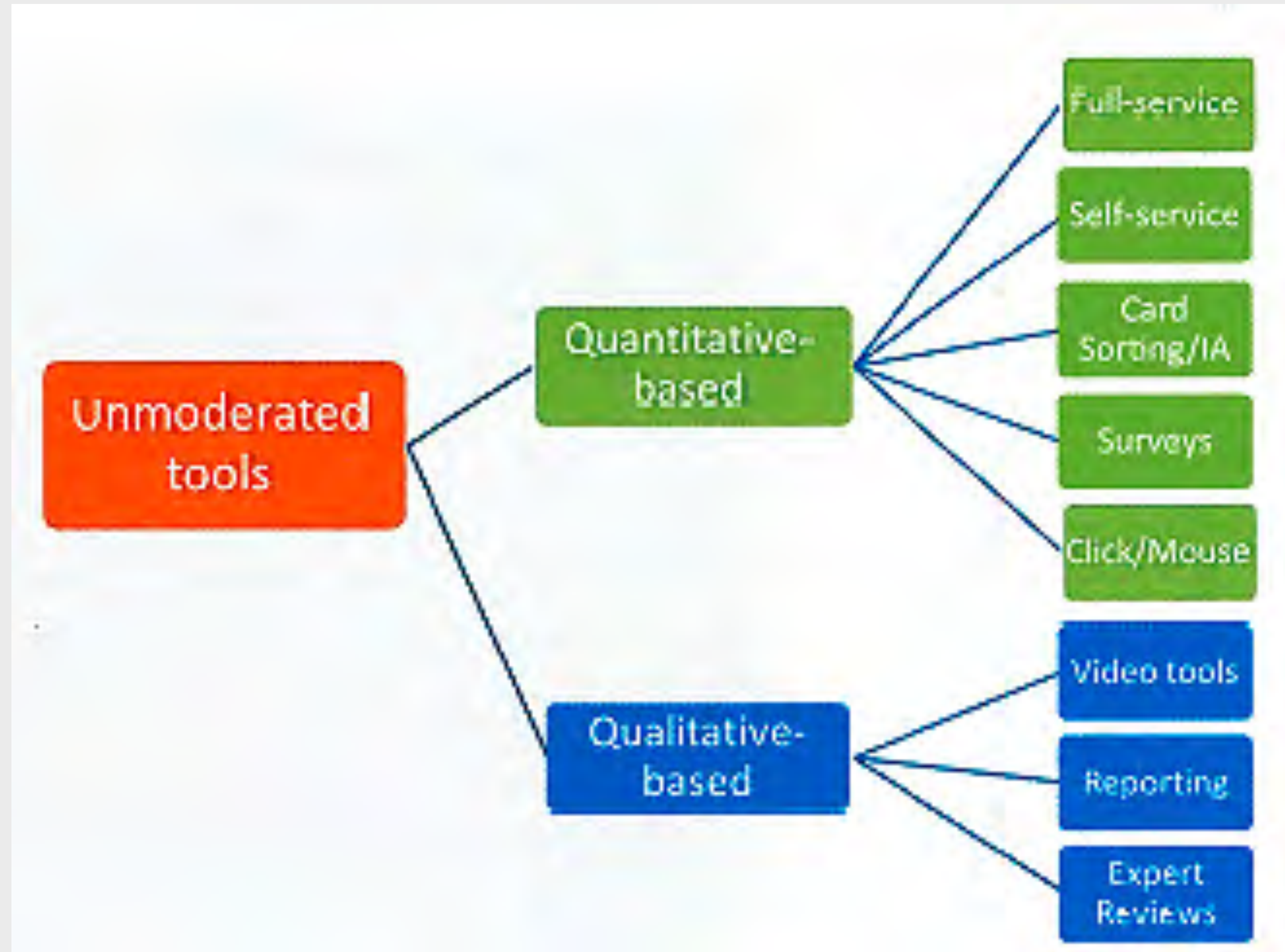
More and more can be done online as technology progresses

## Full Service Tools

- WebEffective
- Self-Service
  - UserZoom
- Card sorting, etc
  - Optimal Workshop
- Surveys
  - Survey Monkey
  - Qualtrics
  - SurveyGizmo
- Click/Mouse
  - Chalkmark, etc

## Qualitative

- [UserTesting.com](http://UserTesting.com)
- Concept Feedback





# Online vs Lab

## Mix and Match based on the question to be answered

### Lab first, then online

- Identify/fix “easy to change” and then focus on remaining tasks with large sample size
- Generate new concepts, ideas, or questions through lab testing and then test/validate online
- Validate attitudes/preferences observed in lab testing

### Online first, then lab

- Identify the most significant issues online through metrics and then use lab study to gather deeper qualitative understanding of those issues
- Collect video clips or more quotes of users to help bring metrics to life
- Gather all the metrics to validate design
  - If it tests well, then no need to bring users into the lab

# Online vs Lab

## A Brief word about surveys....

### Internal Validity

- Does it measure what it is supposed to measure and does it do that completely for the target population?
- Does it represent what it intends and claims to represent
- Avoids Confounding issues

### External Validity

- Validity beyond the study
- The generalizability across the entire population and across context

### Reliability

- Stable results across many instances of surveying

# Online vs Lab

## A Brief word about surveys....

### Validate the surveys

1. Clearly identify your object of measurement. Abstract concepts like “sustainability” or “engagement” need to be operationalized in concrete terms.
2. Define the relevant elements of this object. To determine its “domain of content,” review the literature; consult stakeholders; and if possible, gather preliminary qualitative data through focus groups, interviews, or surveys.
3. Develop measures (e.g., survey questions) for each element.
4. Review and pilot test the survey. Depending on available budget and time, you might talk through the survey questions with potential respondents, ask colleagues to review them, and/or select a few potential respondents and ask them to complete the survey and provide feedback on the content.

# Online vs Lab

## A Brief word about surveys....

### Scales

**How many points? (mid point or no midpoint, 5 vs 7 vs 10, alternating + or - questions)**

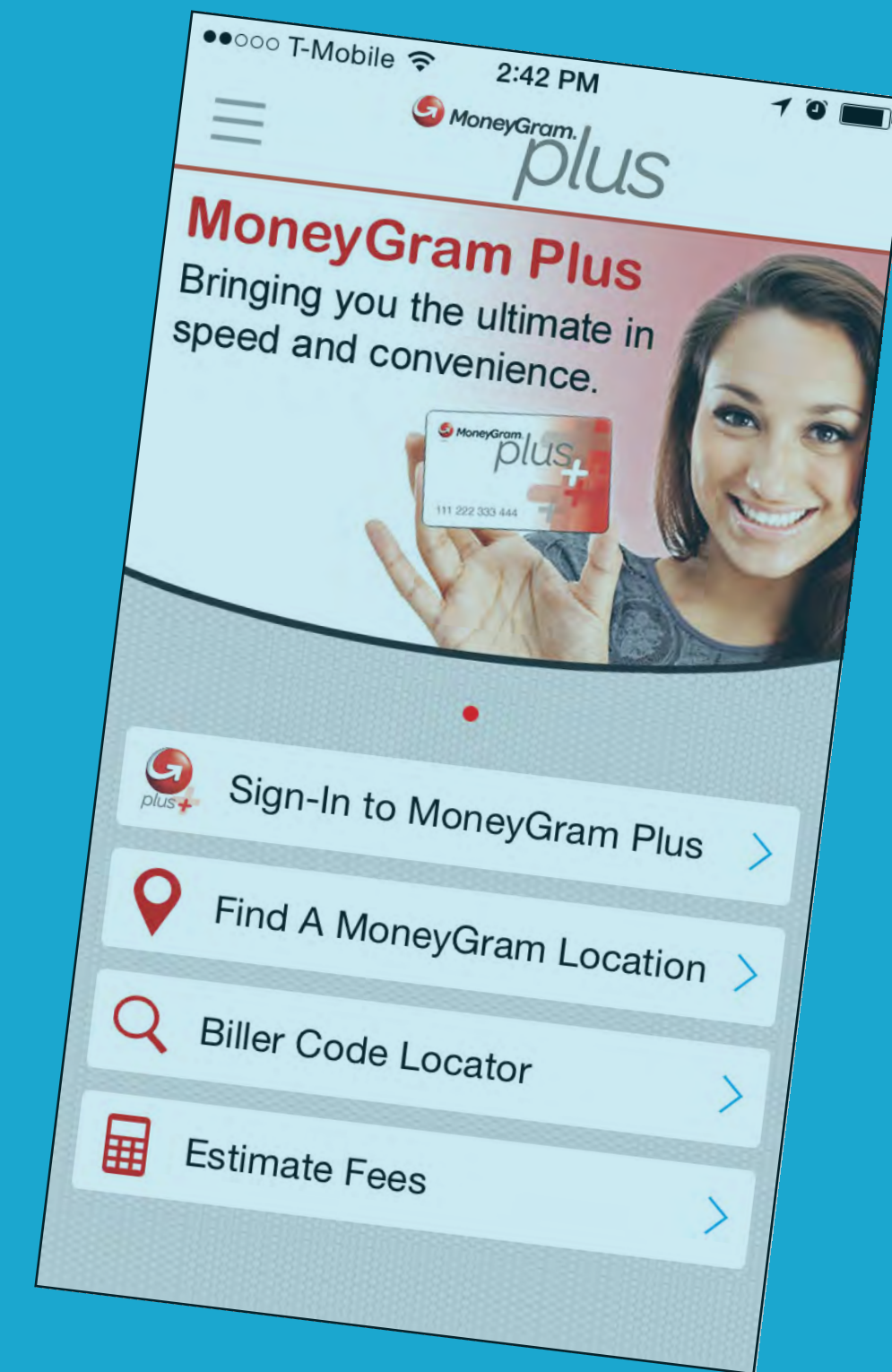
**What about SUS?**

**What is Anchored?**

# Reporting

August 2014

# MoneyGram PLUS UXR Review



## Our Philosophy

To create a visually pleasing and trusting dialog with the user that is easy to navigate, easy to complete with minimal user decision points and interactions that maximizes completion and revenue.

“Never give your user a reason to leave....”

## **Purpose**

**This document was prepared at the request of the client to review and analyze the user experience for MoneyGram PLUS**



# Methods

User Experience Research & Design applied desktop and mobile laboratory validated best practices such as 'first click analysis' and focus group methodology in order pinpoint areas of concern within the MoneyGram Plus App

# Scenario

*“You saw a commercial about an app that would allow you to send money and you thought it would be nice to try it out. You don’t remember too much from the commercial so you really don’t know much about the app other than the commercial said it was fast and easy. You’ve just downloaded this app, installed it, and now you’d like to explore it and try to send money to your mother in Williamston, NC. You’ve never used this application before, but you’d like to use it to perform a series of tasks that we’ll talk about in just a few minutes. ”*

# Tasks

**Task 1:** Assume that you've installed the App on the iPhone and you want to check it out. Take a few minutes to look at the main screen and use the talk out loud method to talk about what you see and your general impressions.

**Task 2:** You want to find out about all the capabilities of this phone app. Locate and tell us all the functions you can perform with this app.

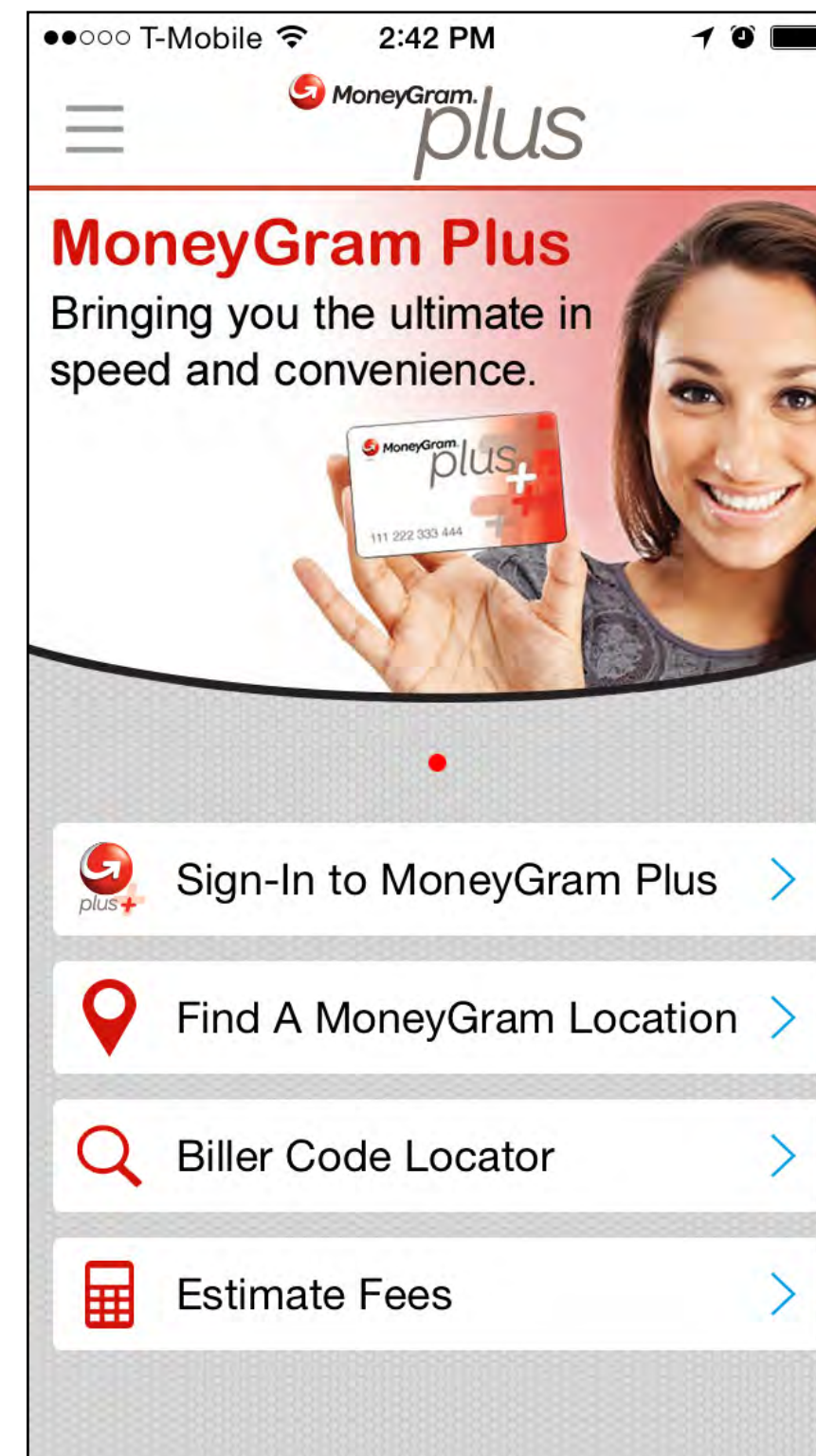
**Task 3:** You've decided that you WILL send money to your mother in Williamston, NC. Find out if there is a place where she can receive the money you want to send to her.

**Task 4:** Send money to your mother in Williamston, NC.

# 1. Task - Overall Impressions

## General Comments

- *"I am already in your app, why are you advertising it?"*
- *"Logo seems overdone"*
- *"Synopsis (in graphic), really doesn't help"*
- *"Picture does not match description"*
- *"Looks like an ad for CVS (drugstore)"*
- *"Feels like a commercial or print ad"*
- *"Is this the home page?"*
- *"I don't trust this app, it looks cheap"*
- *"The 'red' is making me anxious"*
- *"Does not look trustworthy, does not look secure"*



## Verbal Descriptors

- *Sketchy*
- *Suspicious*
- *Confused*
- *Inconsistent*
- *Confusing*
- *Neutral*
- *Uninformative*
- *Untrustworthy*
- *Uninspiring*

## Perceptual Rating (2.75 out of 7.0)

On a scale of 1 to 7, where 1 is not at all attractive, and 7 is very attractive, how would rate the attractiveness of this app?

## 2. Task - What can I do with this app?

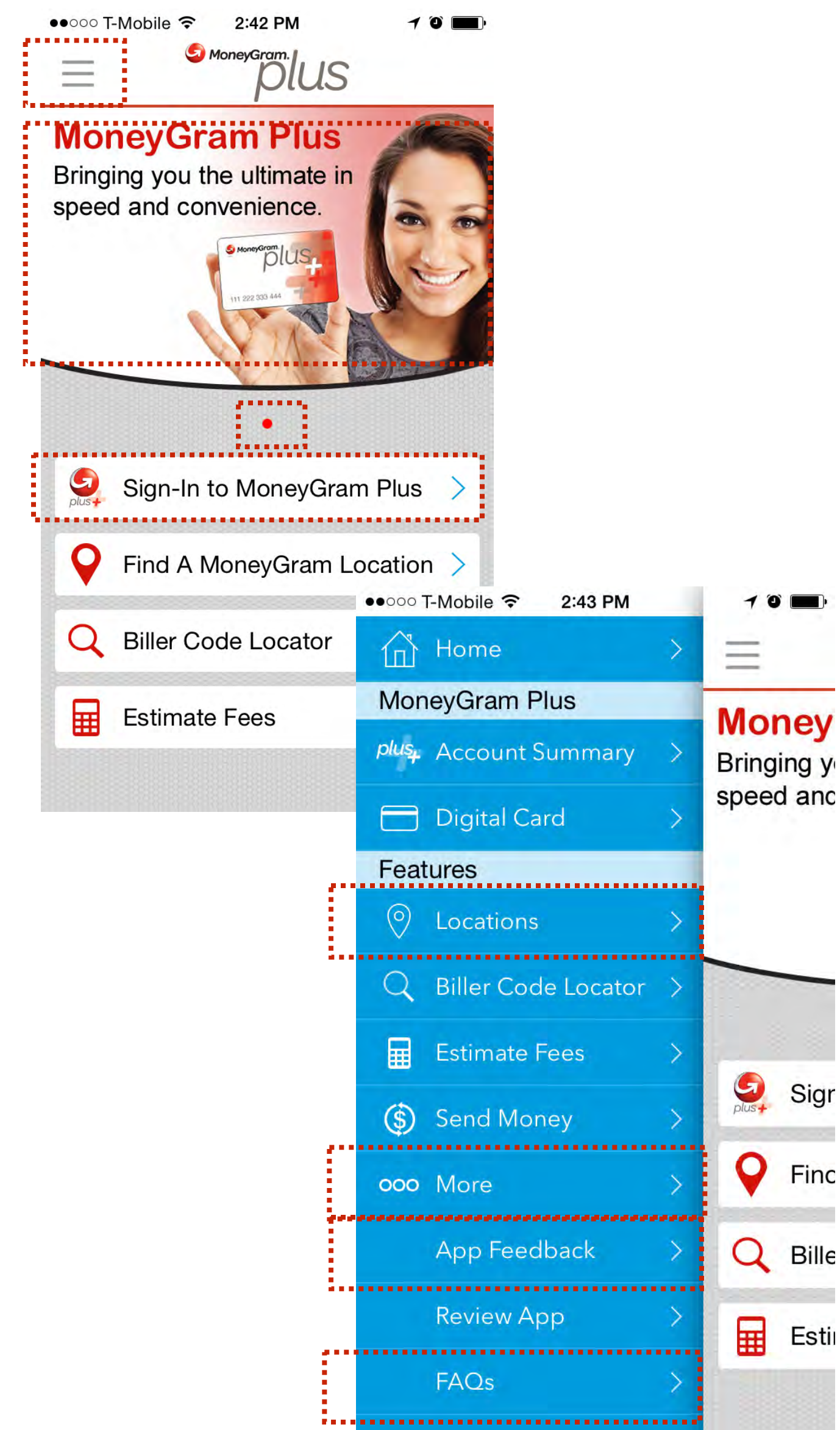
### First Click Analysis\*

“Where would you go to find out all the things this app would be able to help you with?”

- Hamburger Menu: 50%
- Home Page Image: 25%
- Red dot below the Home Page Image: 12.5%
- Sign In to MoneyGram Plus: 12.5%

### Second Click Analysis:

- Of the people who chose Hamburger menu, no one chose **‘Locations’** to find out all the ways to send money and pay bills.
- From the Hamburger menu, some users clicked on ‘More’,
- Some users clicked on ‘Review App’
- FAQs.



\*Jeff Sauro of Measuring Usability cites research supporting the importance of First Click Testing.

A participant who clicks down the right path on the first click will complete their task successfully 87% of the time.

A participant who clicks down the wrong-path on the first click, tends to only successfully complete their task 46% of the time.

# 3. Task - Where Can Money Be Received?

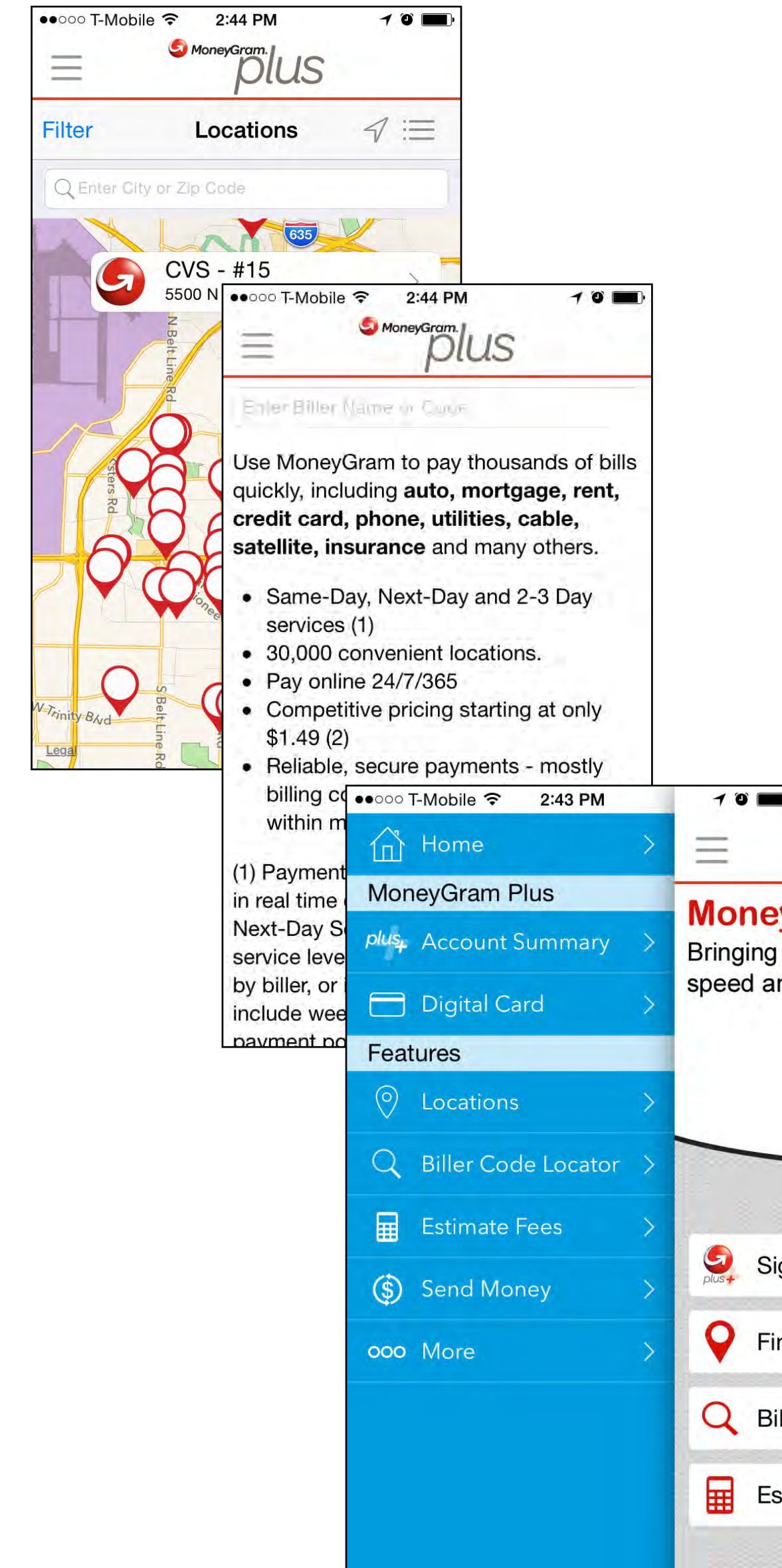
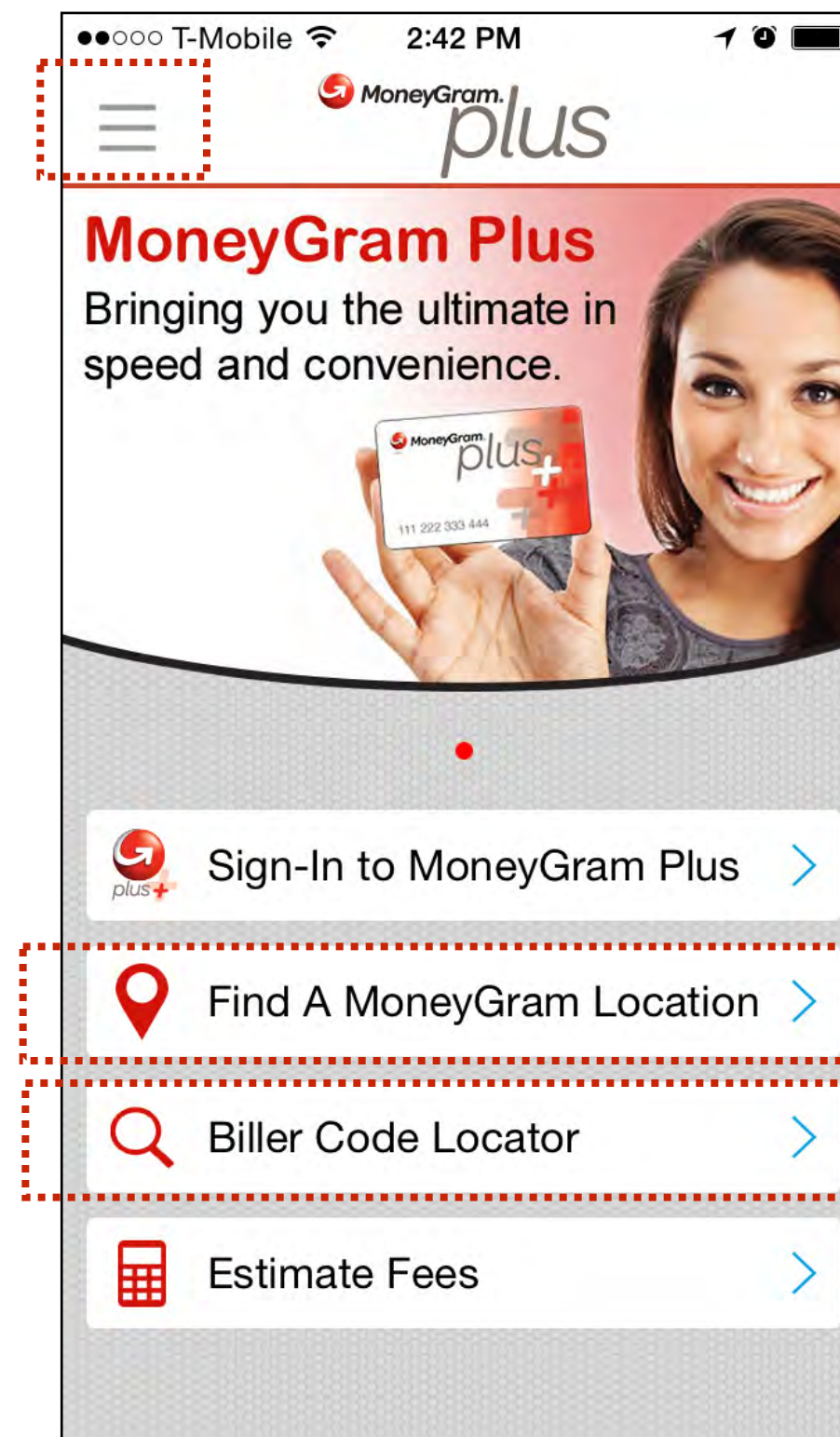
## First Click Analysis

“Where would you go in order to find out if your recipient had a location near them?”

- Find a Moneygram Location: 62.5%
- Biller Code Location: 25%
- Hamburger Menu: 12.5%

## Observations:

- The magnifying glass icon (Biller Code Locator) was misleading. 25% of the people that it was ‘Search’.



\*Jeff Sauro of Measuring Usability cites research supporting the importance of First Click Testing.

A participant who clicks down the right path on the first click will complete their task successfully 87% of the time.

A participant who clicks down the wrong-path on the first click, tends to only successfully complete their task 46% of the time.

# 4. Task - Sending Money

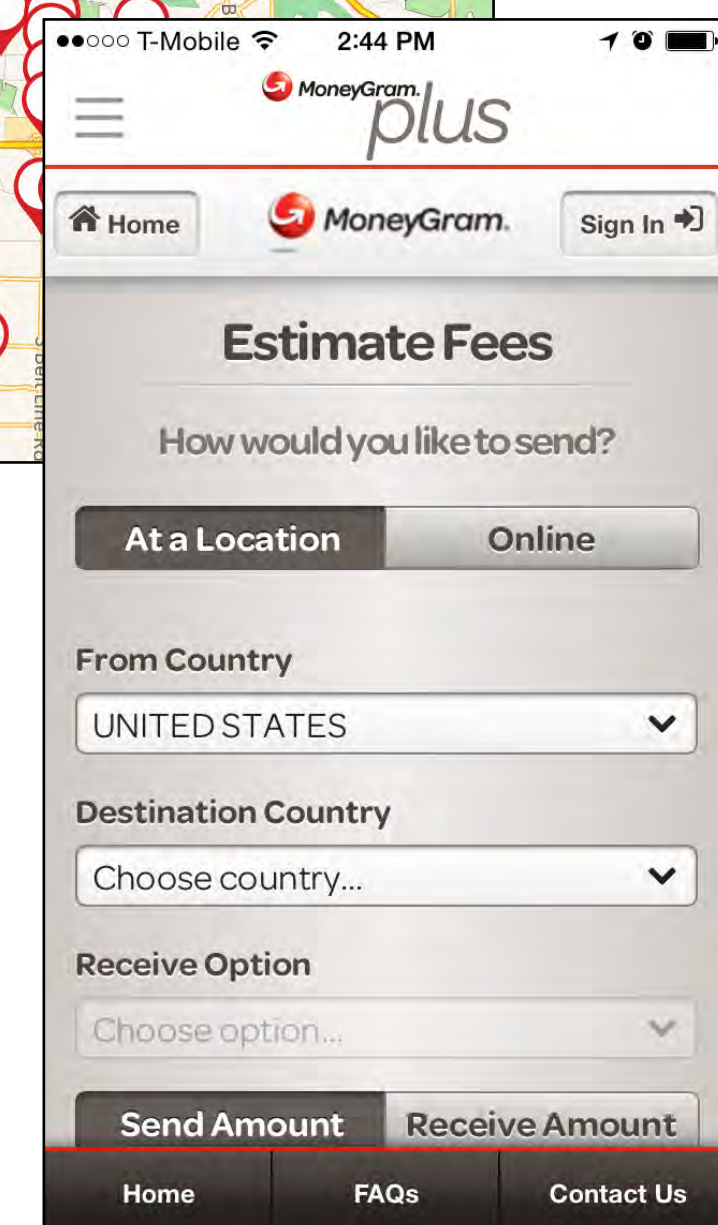
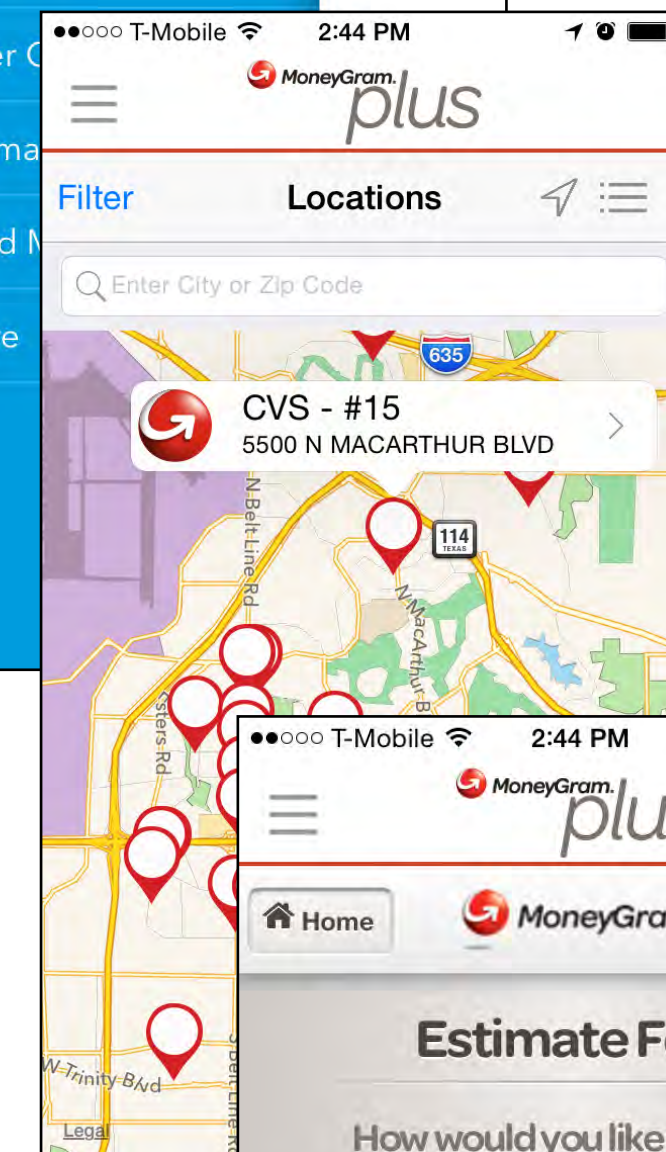
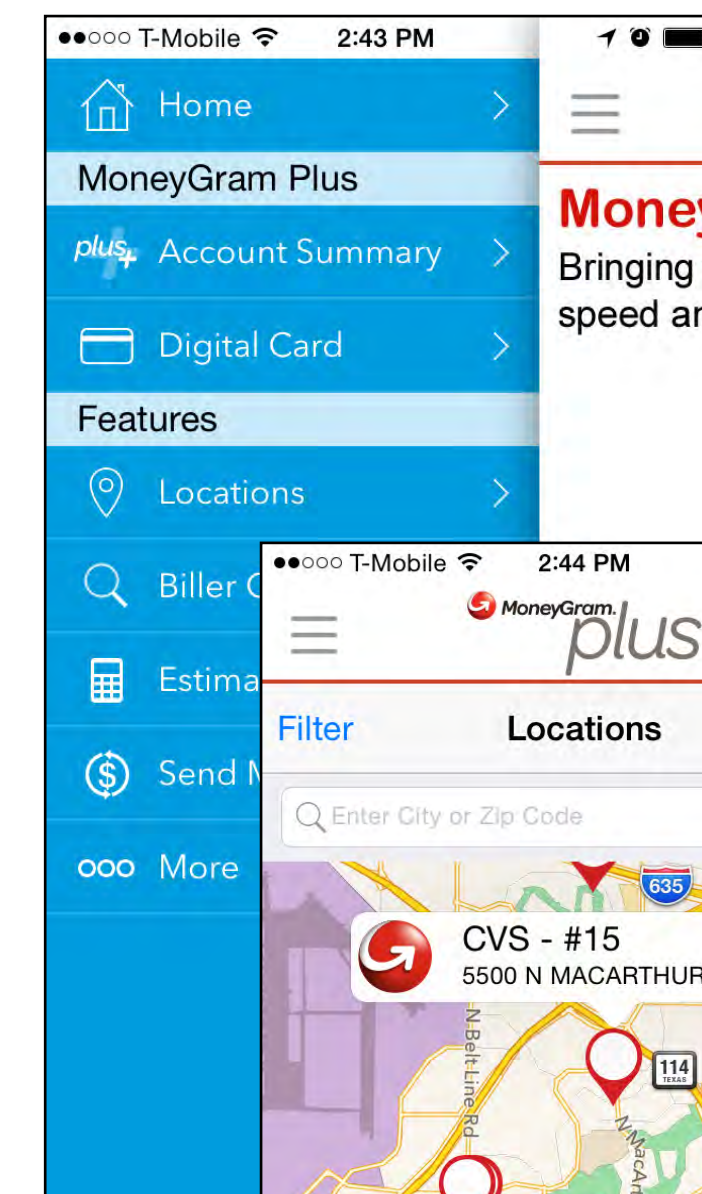
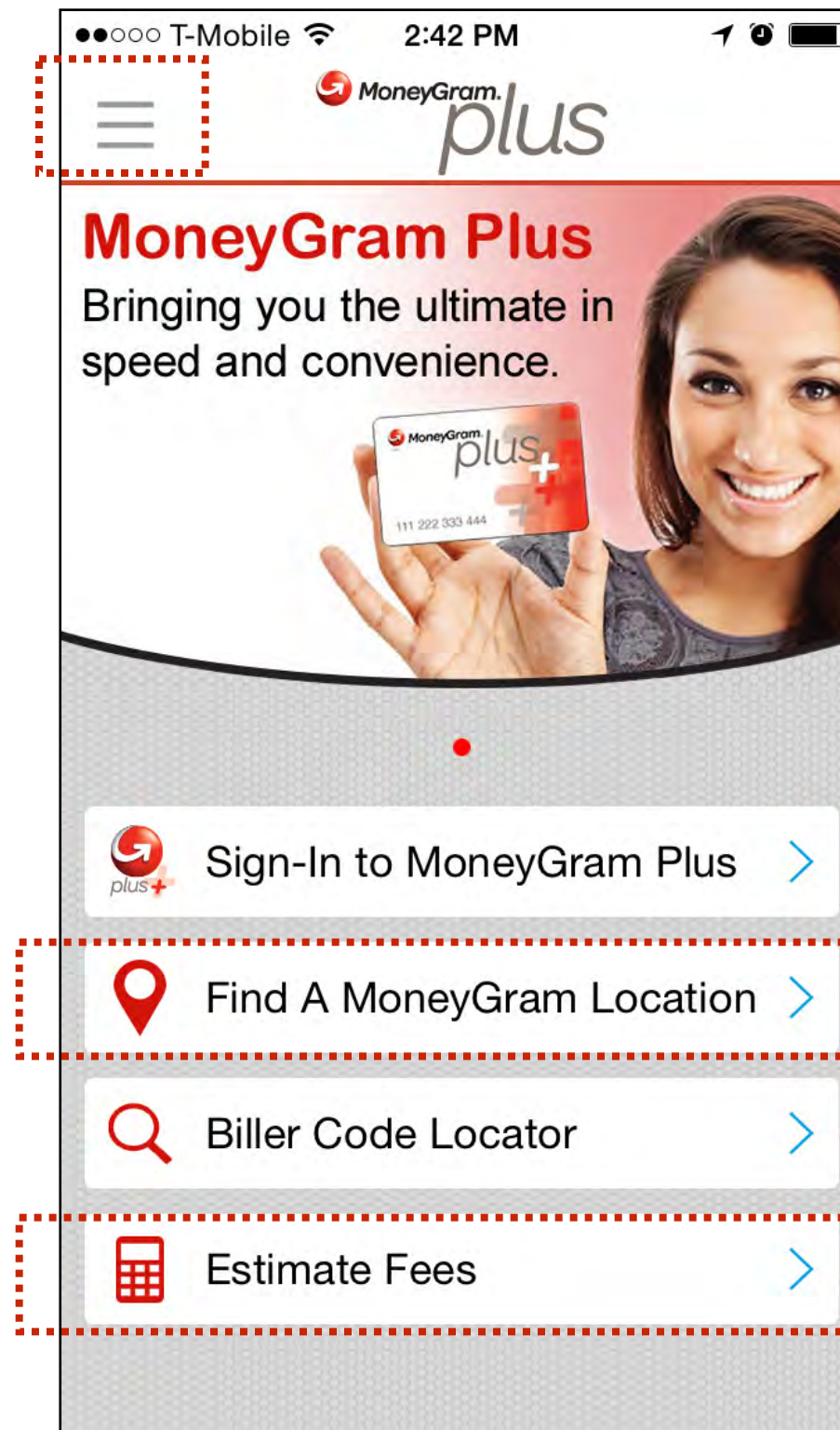
## First Click Analysis\*

“Send money to your recipient.”

- Hamburger Menu: 50%
- Find a Moneygram Location: 37.5%
- Estimate Fees: 12.5%

## Comments:

- “I wasn't sure where to click”
- “I thought i could send the money by selecting location (from the map)”
- “I felt stuck”



\*Jeff Sauro of Measuring Usability cites research supporting the importance of First Click Testing.

A participant who clicks down the right path on the first click will complete their task successfully 87% of the time.

A participant who clicks down the wrong-path on the first click, tends to only successfully complete their task 46% of the time.

# Post Task Comments

## TRUST / INFORMATION

- *After not finding information “about the app”, I would have abandoned (this).*
- *I Don't trust apps or anything that deals with money and doesn't explain (itself).*
- *I Don't trust red color.*
- *The app is so confusing and the way its designed makes it look cheap. I don't feel like I could trust it.*
- *The color scheme is confusing- the menu is blue and everything else is red. Why?*
- *The overall feels of the app is very 'off-brand' looking. The colors make the feel very discomforted along with the look. Its already nerve-wrecking to send money transfers outside of my official bank.*
- *The app is so confusing and the way its designed makes it look cheap. I don't feel like I could trust it.*
- *I would have abandoned it almost right away (looks cheesy and fake). when it comes to money, I only use trustworthy interfaces. If it is at all confusing and unprofessional, i am out.*
- *The word 'Moneygram" repeated 5 times. I am aware I am in the money gram app, so why repeat 5 times?"*

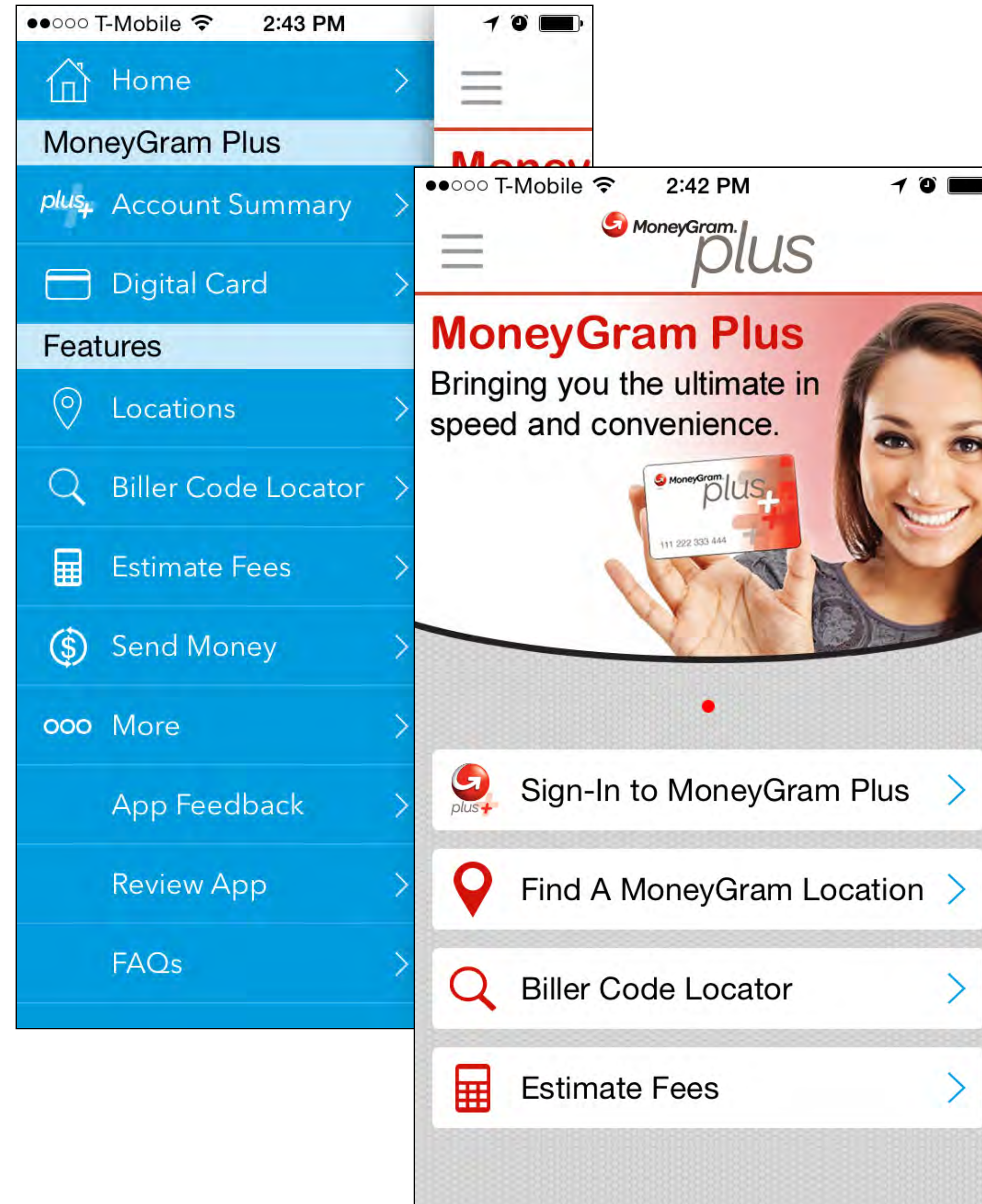




# Post Task Comments

## TASK ORIENTATION

- Overall my feeling is really confused. the icons and names of menus don't match.
- Not getting the information I am looking for, makes me want to leave the app.
- 'Biller code locator' - text is unclear- billing who, billing what, what code?
- Don't understand some of the buttons- terribly named. So many 'Go home' buttons.



# Post Task Comments

## Signing In

- Upon entering I might abandon just because of the need to sign in (when I don't have an account). I would definitely abandon upon trying to sign up and having to go to the web page.
- Can't make an account- its is frustrating! and it leads me to desktop page?

