

# Human Computer Interaction CS449 – CS549

Week 11

---

## Measuring UX Methods

KÜRŞAT ÇAĞILTAY

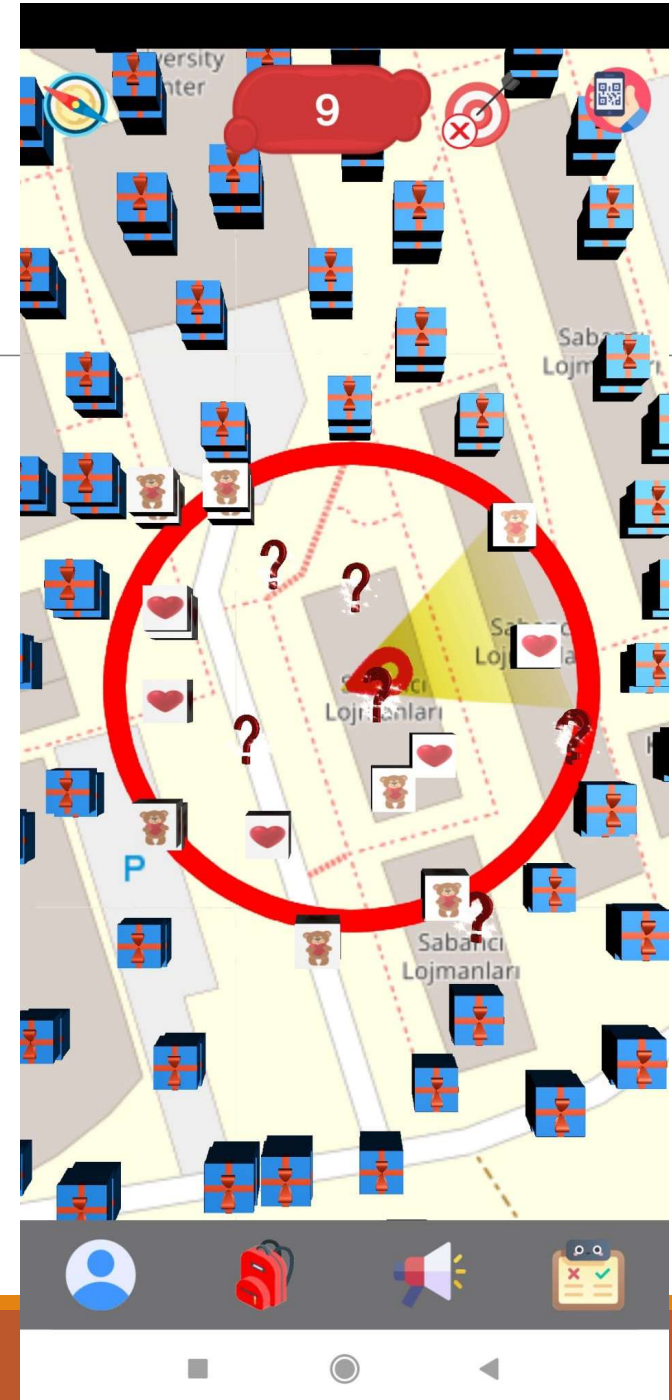
# Today

---

- Measuring UX Methods
- Term Project Report

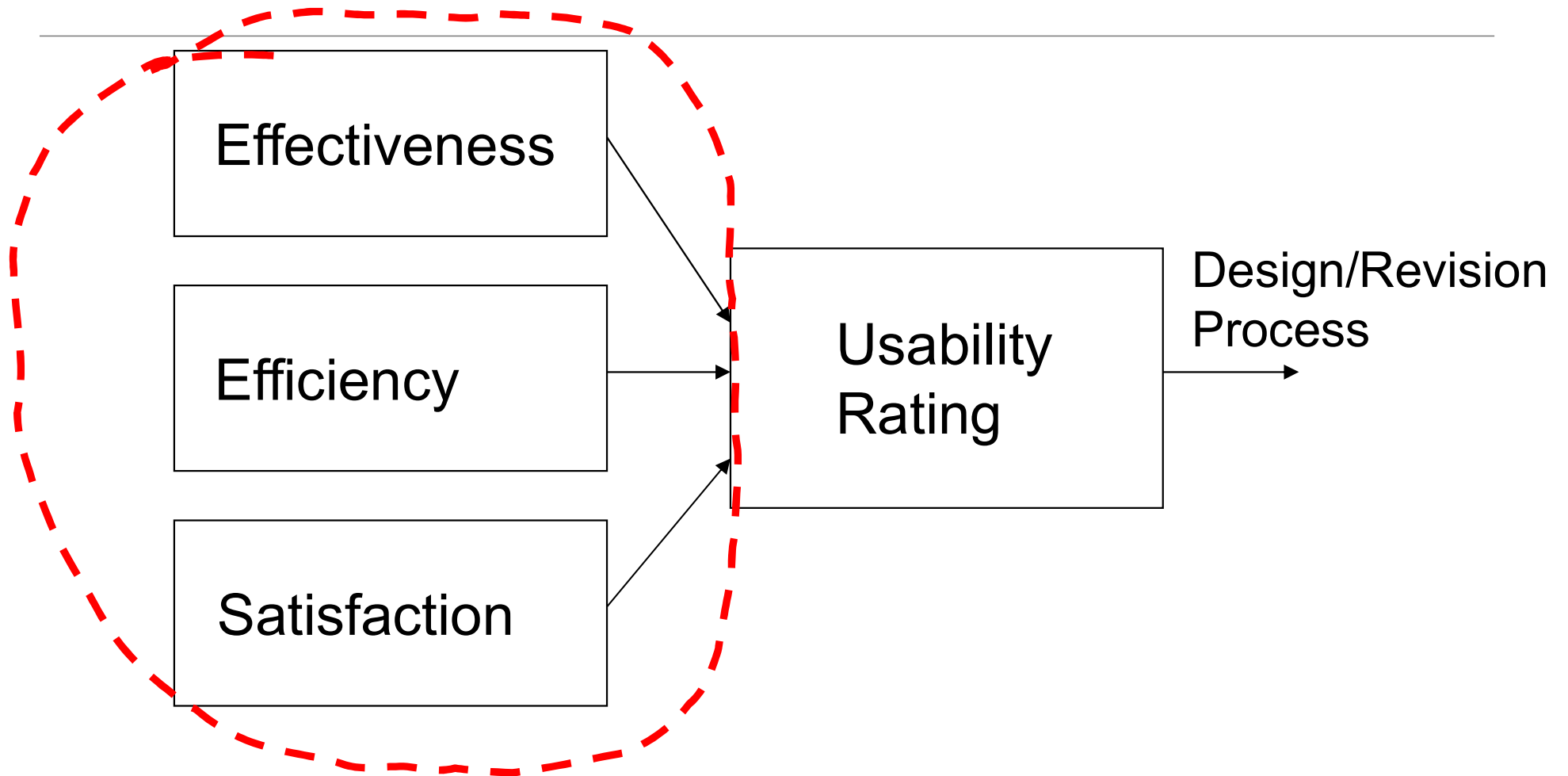


# Top scorers?

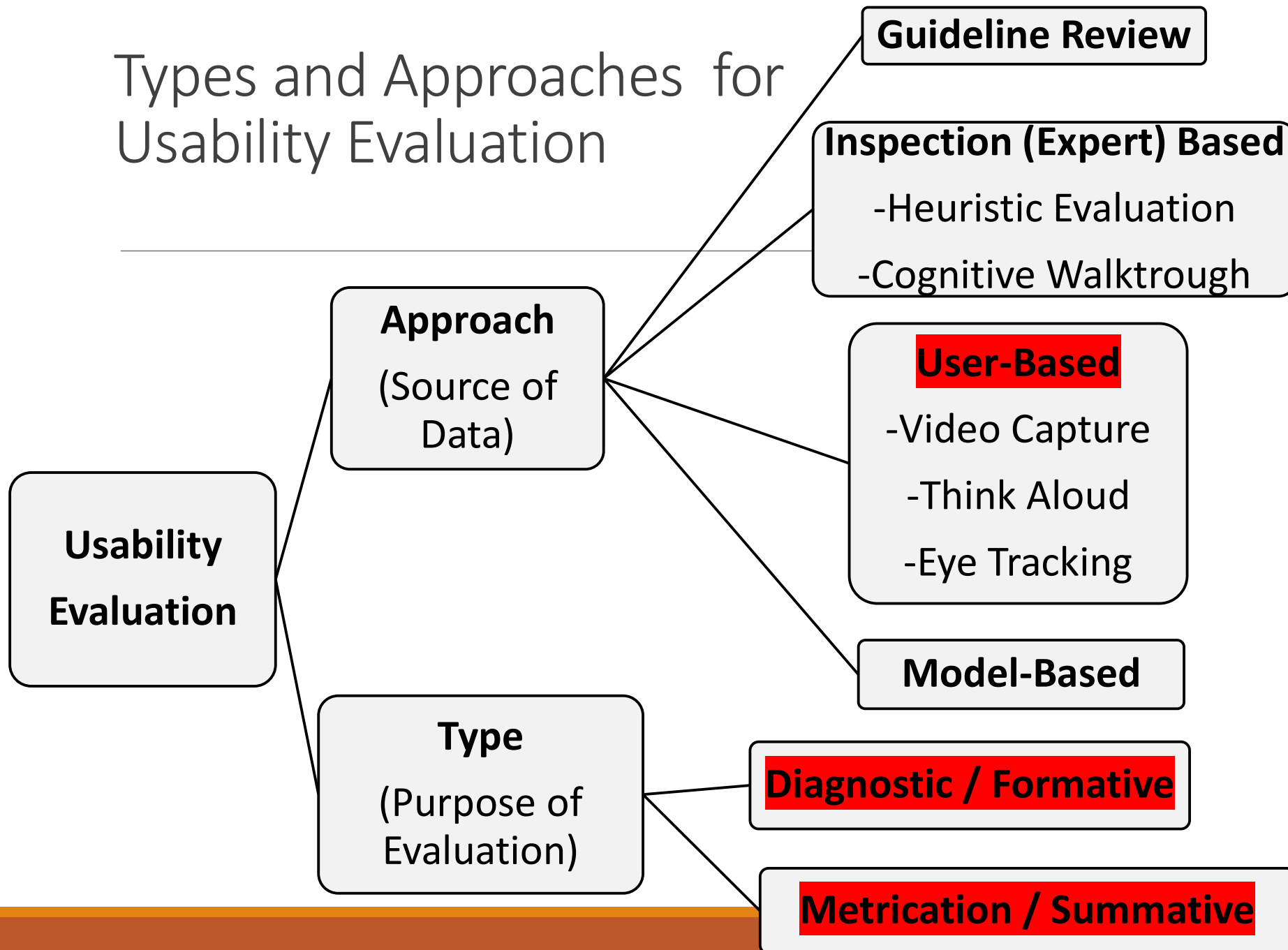




# Determinants of usability rating



# Types and Approaches for Usability Evaluation



# Evaluation Type

Formative Evaluation

Summative  
Evaluation



Project timeline: Analysis, Design... >>>

Final product



## User Based Testing

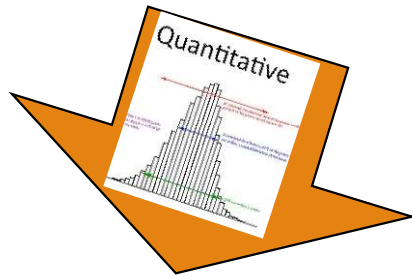
# User Based Testing





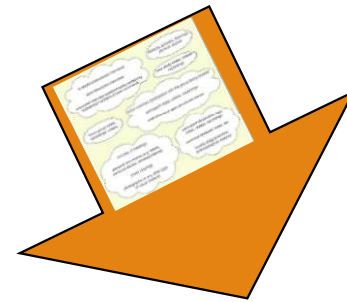
# Research Methods: Data Collection

---



## **Quantitative Methods (Statistics)**

- Efficiency
- Effectiveness
- Satisfaction



## **Qualitative Methods (Verbal - Mostly)**

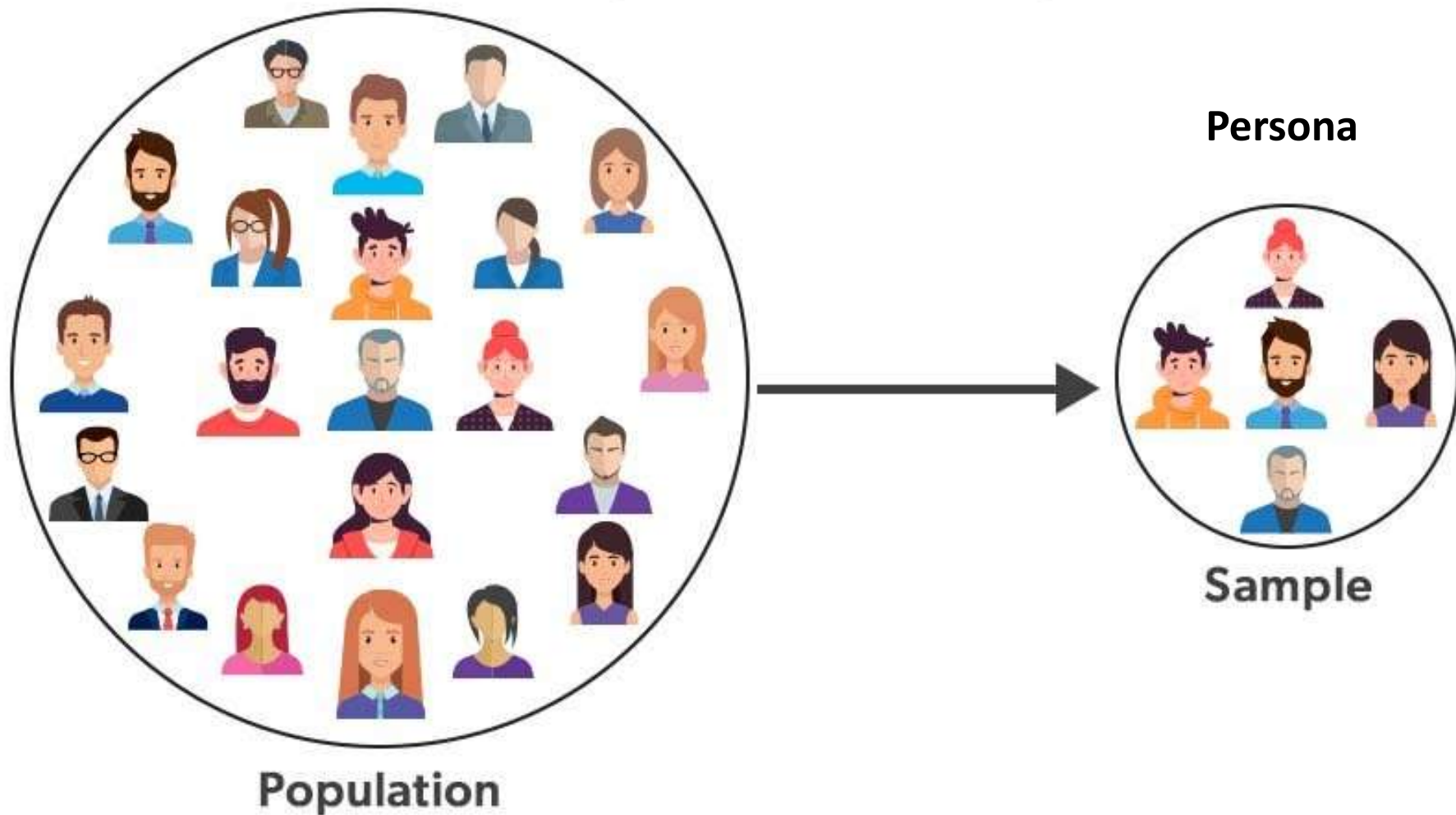
- Satisfaction

# Empirical Studies

---

- To understand **cause** and **effect**
  - “When small size buttons are used, it decreases likelihood of completing a purchase”
- To make **predictions**
  - “When users type on a new keyboard, their typing speed increases.”
- To test **hypotheses**
  - “There is no performance difference between reading from paper and screen.”

## Population and Sample



Hypothesis: *“My new keyboard is easy to use”*

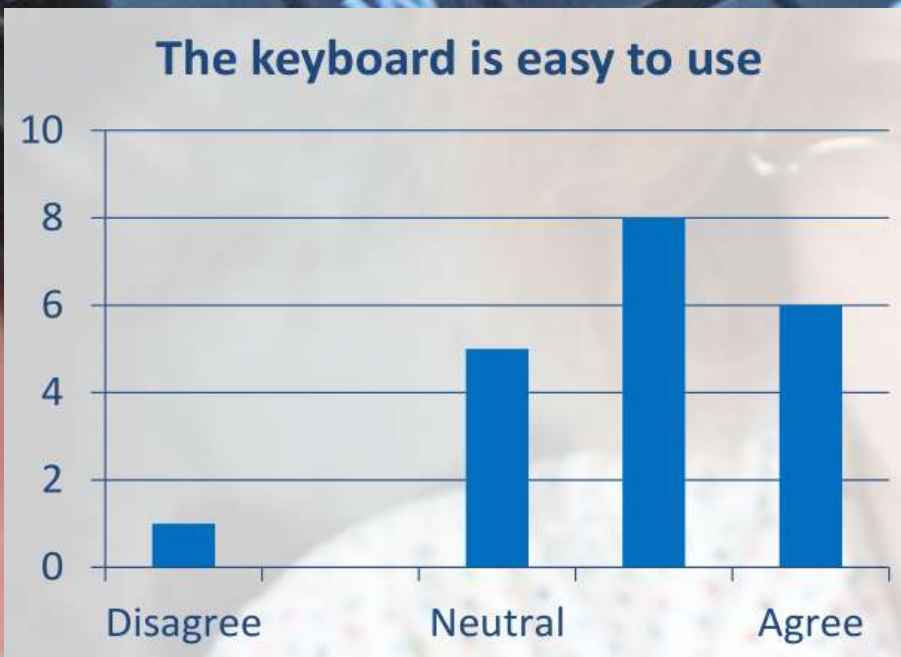


Photo by Niels Henze

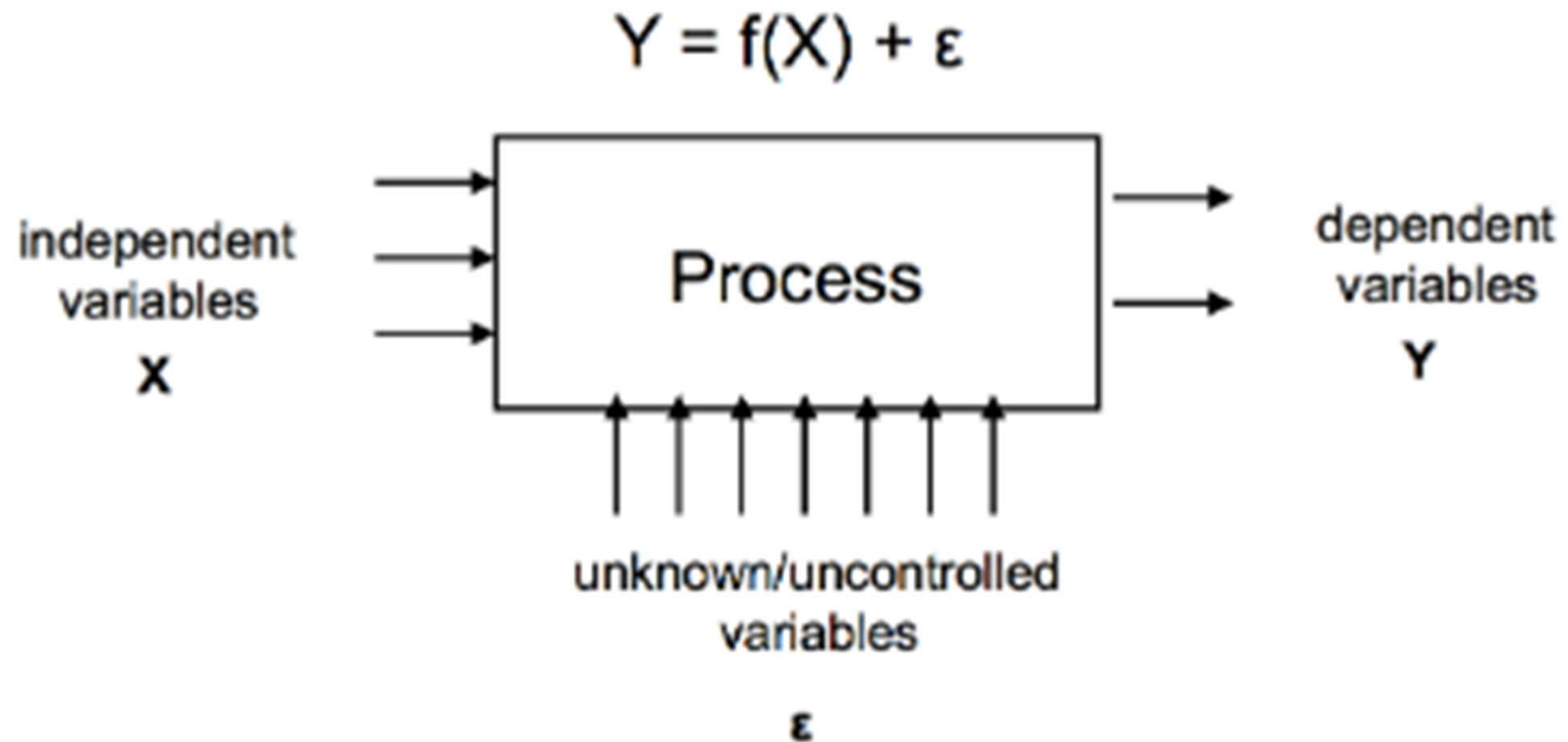
# Observation vs Controlled Experiments?

---

- Participants rated the system easy to use, because
  - they actually find the system easy to use?
  - they want to make you happy in your research?
- Knowing the reason for our observation helps us predict things about the world
  - But a mere observation will not help to find the answer!

# Experiment Design

---



# Controlled Experiments

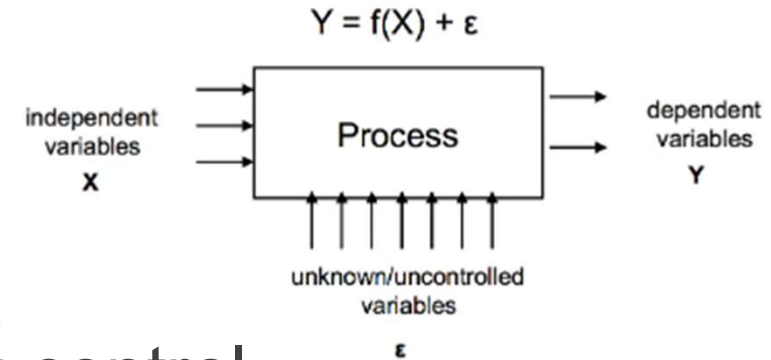
- Controlled experiments are means to isolate cause and effect



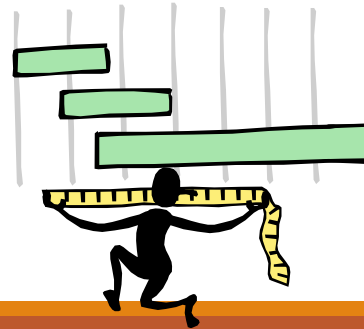
- What if there are potential two effects or if they potentially depend on each other?



# Experimental Variables



- Independent Variables (IV) : the ones you control
  - Aspects of the interface design (e.g. Color, Menu size)
  - Characteristics of the testers (e.g. Male/Female, )
  - Continuous: Time between clicks for double-click
- Dependent Variables (DV) : the ones you measure
  - Efficiency & Effectiveness
  - Time to complete tasks
  - Number of errors
  - Satisfaction Scores





# Descriptive vs. Inferential

---

- Descriptive statistics
  - Summarize a group of numbers from a research
- Inferential statistics
  - Draw conclusions/make inferences that go beyond the numbers from a research study
  - Determine if a causal relationship exists between the Independent and Dependent Variables

# Data analysis and interpretation

---

## **descriptive**

mean

median

mode

variance

standard deviation

## **inferential**

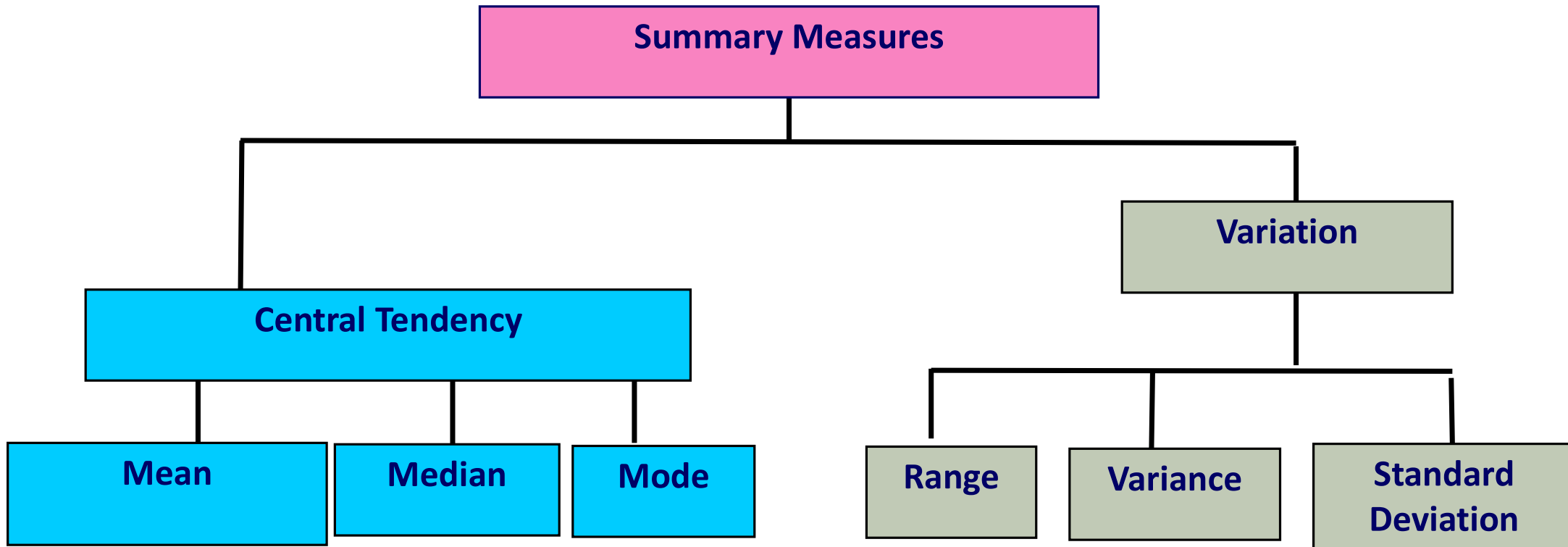
t-test

analysis of variance

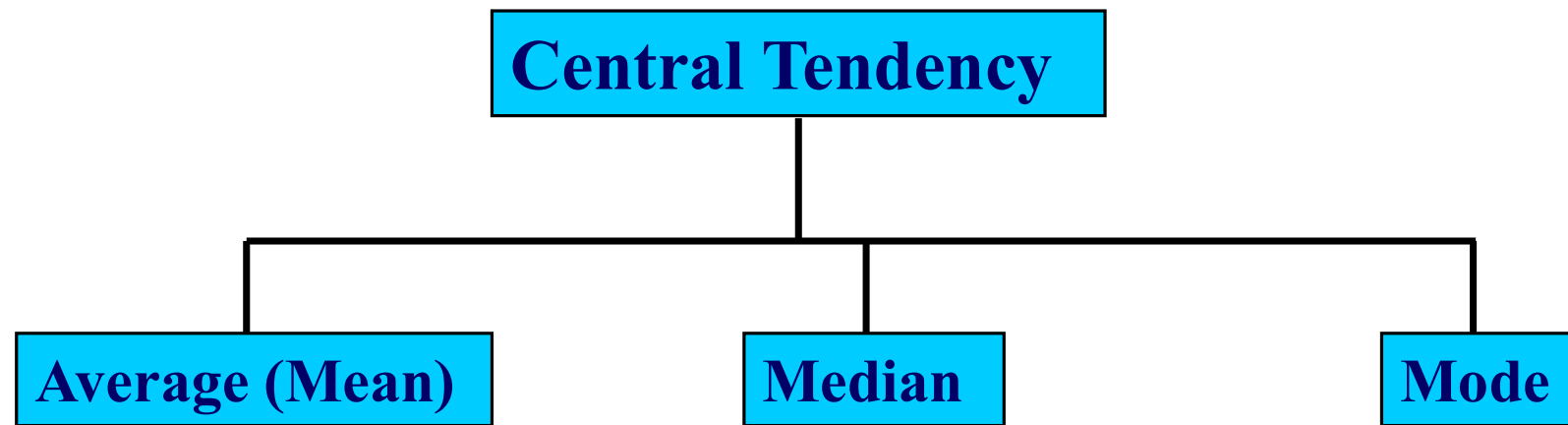
chi squared

# Summary Measures

---

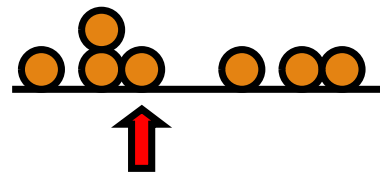


# Measures of Central Tendency



$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

$$\mu = \frac{\sum_{i=1}^N X_i}{N}$$



**mode**

The mode is the most frequent value.

- Count how many of each value appears.
- The mode is the value that appears the most.
- You can have more than one mode.

2, 2, 3, 5, 5, 7, 8

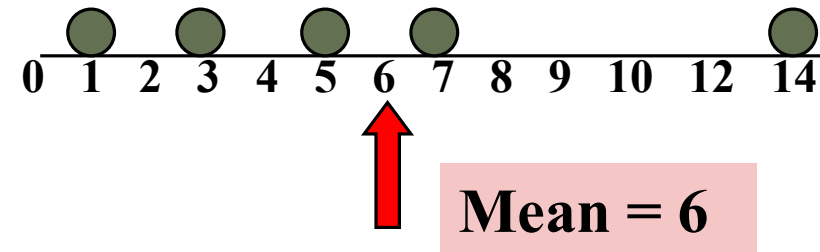
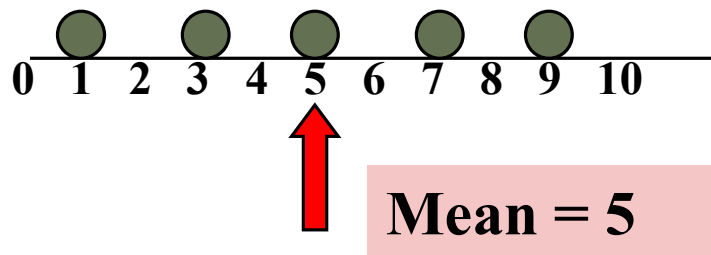
**2** **5**

The modes are 2 and 5

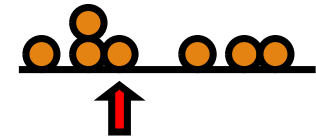
# Mean

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

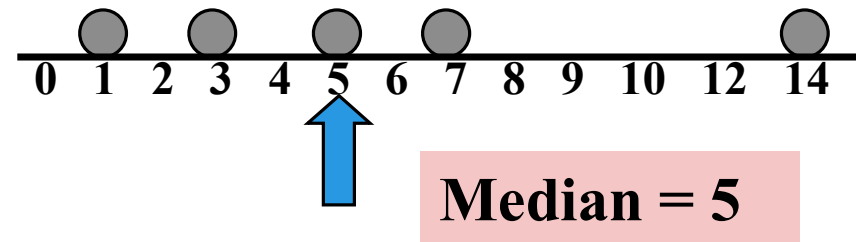
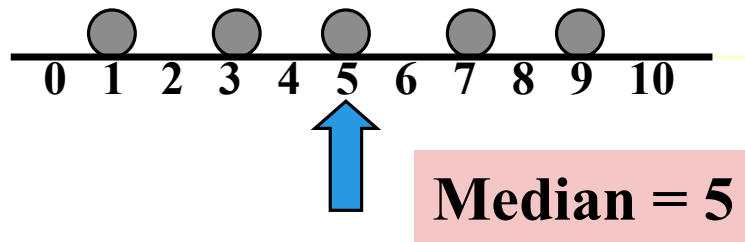
- The most common measure of central tendency  $\mu = \frac{\sum_{i=1}^N X_i}{N}$
- E.g. Users completed tasks in 2 minutes average
- Affected by extreme values (outliers)



# Median - Middle

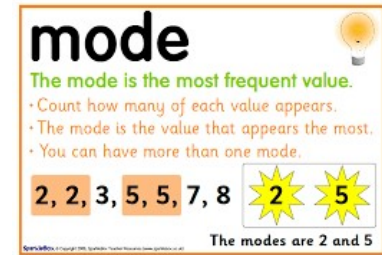


- Robust measure of central tendency
- **Not** affected by extreme values

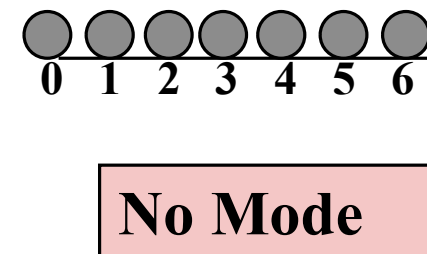
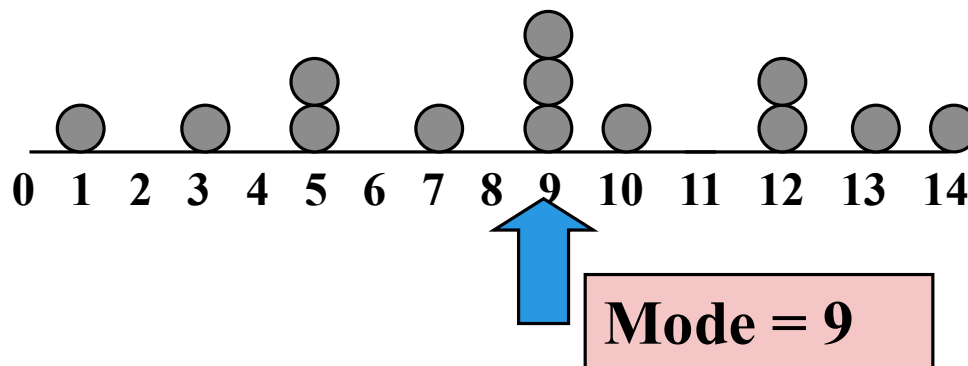


- In an **Ordered array**, median is the “middle” number

# Mode

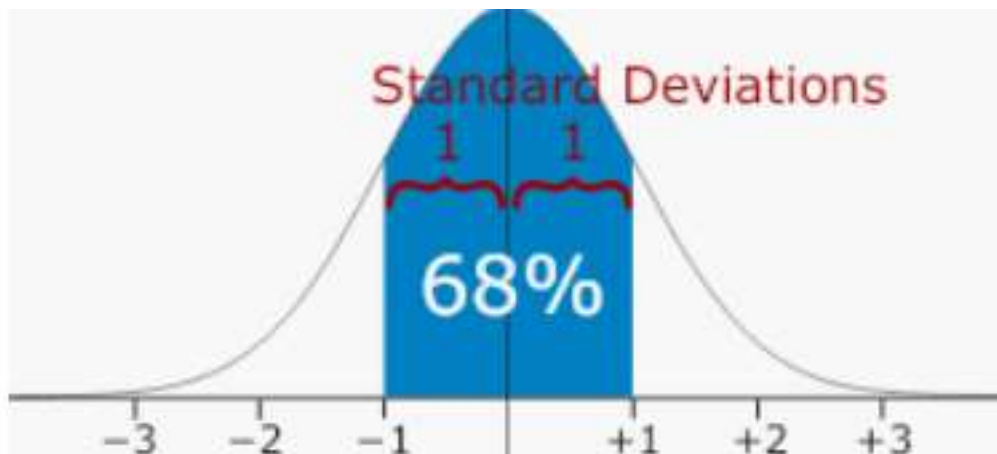


- A measure of central tendency
- Value that occurs most often
- Not affected by extreme values
- Used for either numerical or categorical data
- There may be no/several mode (s)



# Standard Deviation

...the descriptive statistic indicating the spread of a set of scores around the mean



## Standard Deviation

$$\sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

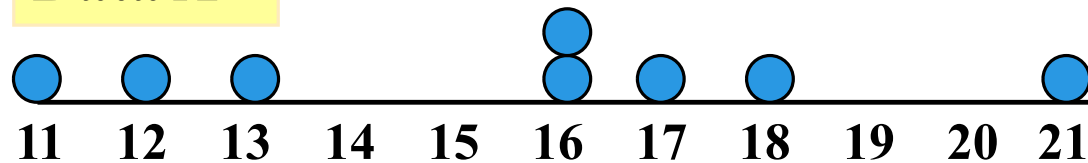
76	84	69	92	58
89	73	97	85	77

$$\bar{X} = \frac{\text{Sum}}{n}$$



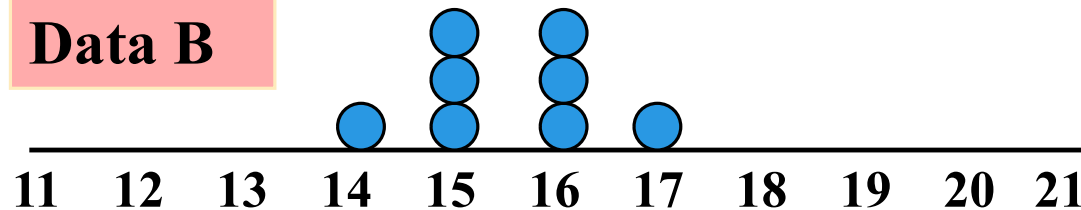
# Comparing Standard Deviations

**Data A**



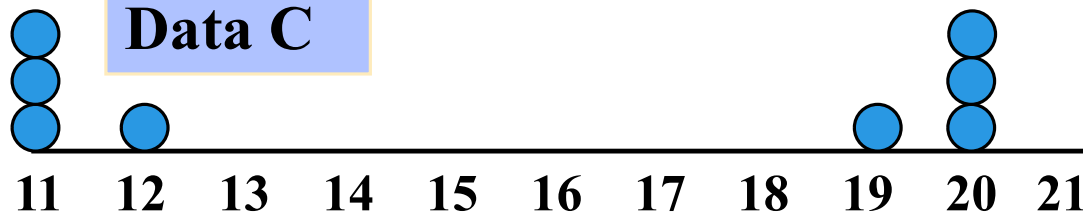
**Mean = 15.5**  
**SD = 3.338**

**Data B**



**Mean = 15.5**  
**SD = .9258**

**Data C**

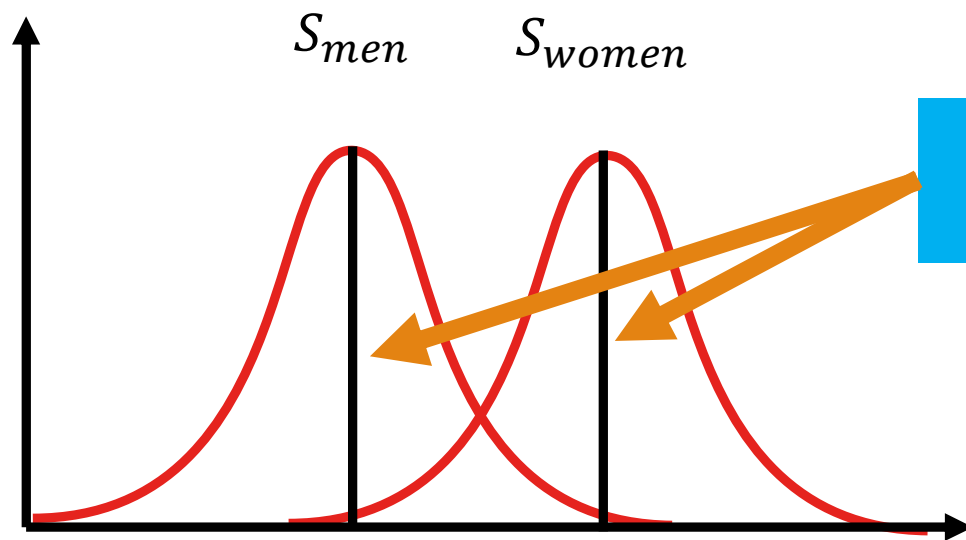


**Mean = 15.5**  
**SD = 4.57**

# Two IVs: Which one is true?

---

- *There is a difference in typing speed between males and females*
- *There is no difference in typing speed between males and females*



difference statistically significant?

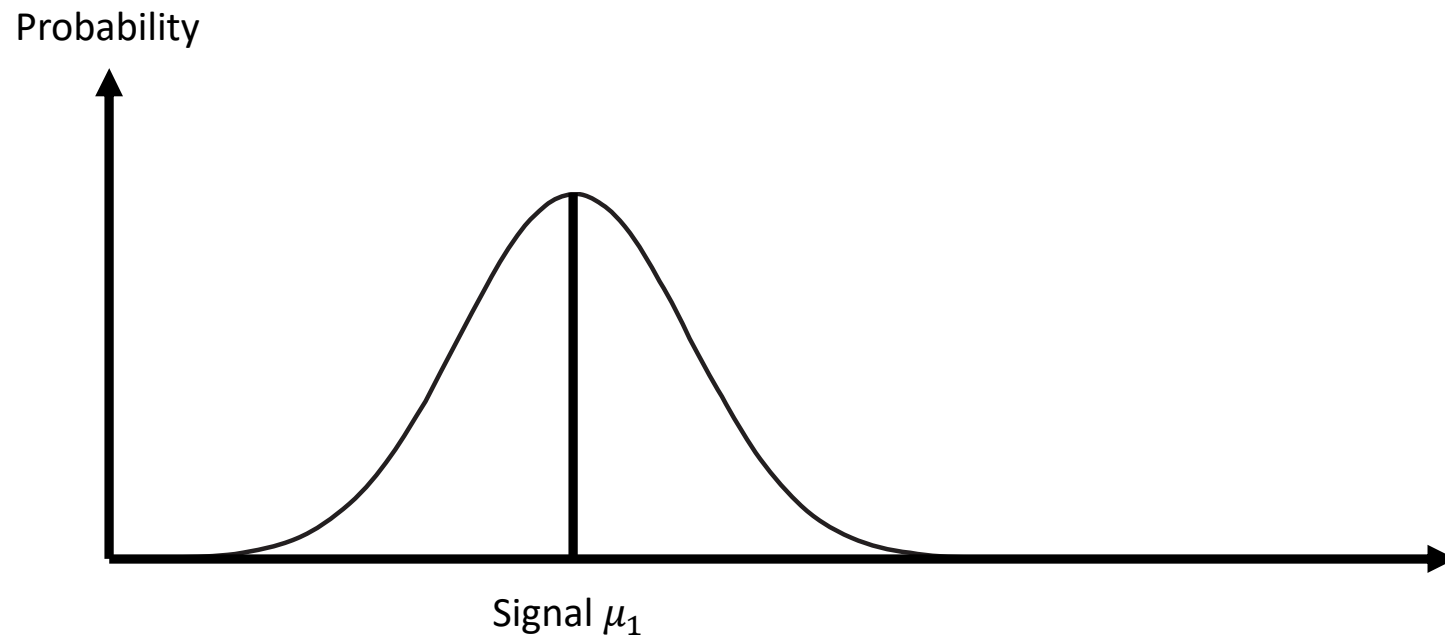
# Comparing Means: T-test

---

- inferential statistic indicating whether the means of *two* groups are significantly different from one another
- Compare means of dependent variable between **two groups**
- How to make t-test with Excel:
- <https://toptipbio.com/t-tests-excel/>
- <https://www.youtube.com/watch?v=q0ckcKsSPXU>

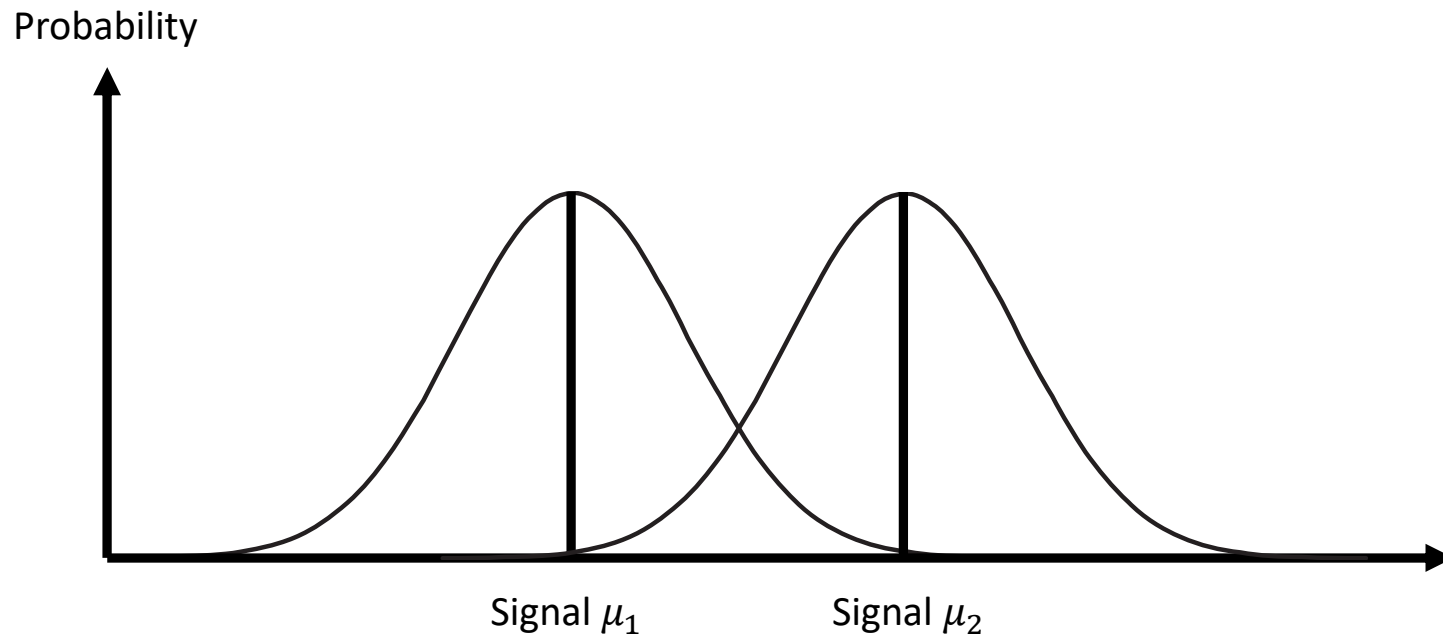
# Hypothesis Testing

---



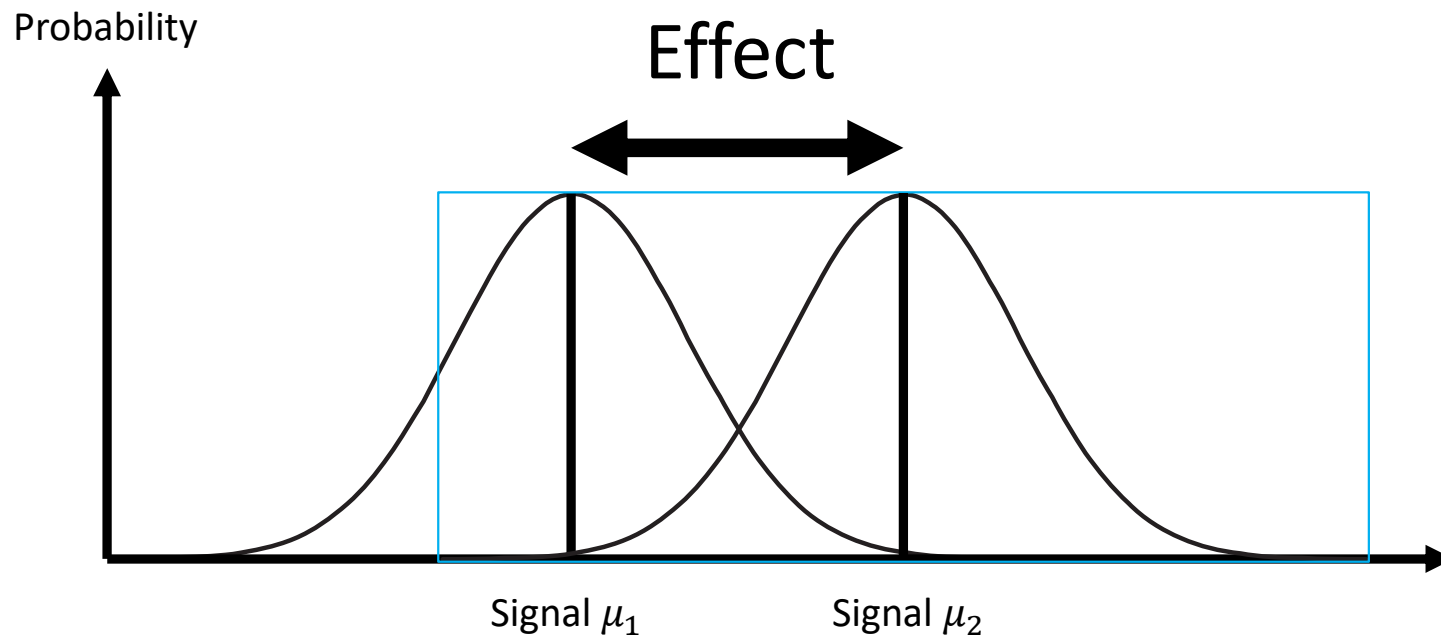
# Hypothesis Testing

---

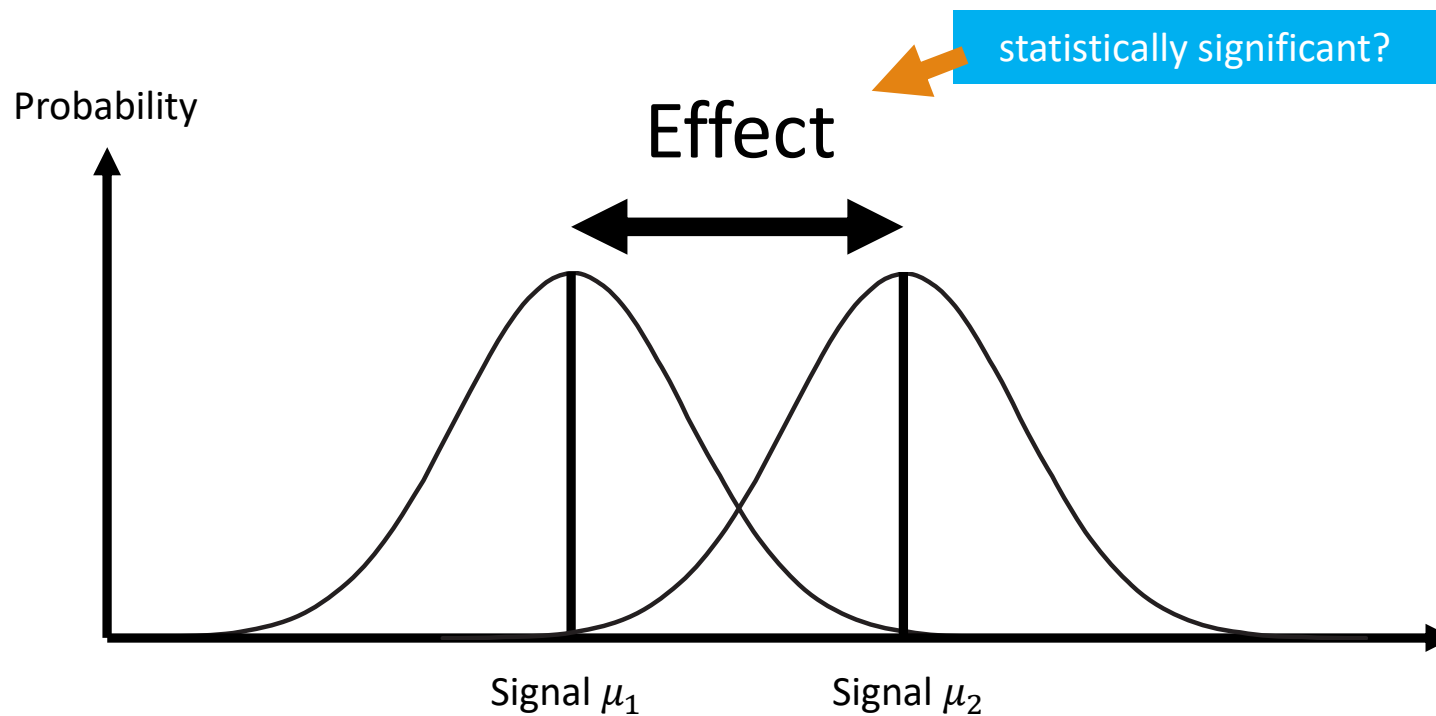


# Hypothesis Testing

---

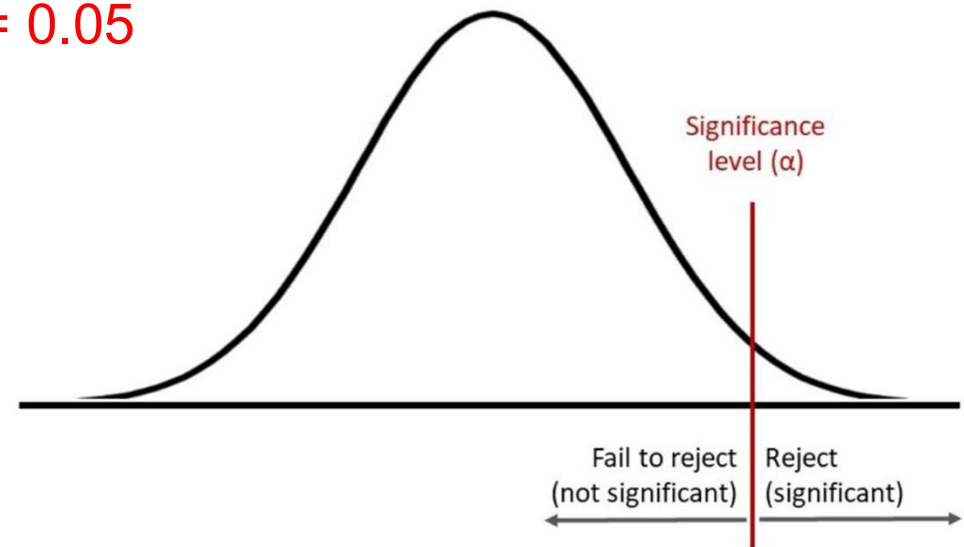


# Hypothesis Testing



# Statistical Significance $\alpha = 0.05$

- A statistical significant effect exists if the probability that the difference occurred is below a certain significance level
- Significance level ( $\alpha$ )
  - Lower significance level means higher evidence
  - Arbitrary, but typical significance level:  $\alpha = 0.05$
- Significant results ( $p < \alpha$ )
  - There is a statistical significant difference
- Non-Significant results ( $p \geq \alpha$ )
  - We cannot conclude anything!





# Example : Gender & Typing CPS

- What can we say from descriptive statistics?



	Male	Female
1	1.89	2.39
2	1.82	1.86
3	7.12	1.82
4	2.30	2.34
5	1.66	1.94
6	1.84	2.01
7	1.80	2.28
8	1.45	2.06
9	1.54	1.91
10	1.72	2.07

Average      2.314      2.068

# Example : Gender & Typing

- If we perform a paired t-test
- $p = 0.67 > \alpha = 0.05$
- No significant difference between the conditions
- We cannot conclude anything



	Male	Female
1	1.89	2.39
2	1.82	1.86
3	7.12	1.82
4	2.30	2.34
5	1.66	1.94
6	1.84	2.01
7	1.80	2.28
8	1.45	2.06
9	1.54	1.91
10	1.72	2.07

Average      2.314      2.068

# Example : Gender & Typing

- Let's assume we draw a different sample
- $p = 0.028 < 0.05$ 
  - Significant difference between the conditions
  - Typing speed results for female higher CPS than male
- One outlier between rejecting and accepting  $H_0$  indicates a weak statistical power!





	Male	Female
1	1.89	2.39
2	1.82	1.86
3	2.30	1.82
4	2.30	2.34
5	1.66	1.94
6	1.84	2.01
7	1.80	2.28
8	1.45	2.06
9	1.54	1.91
10	1.72	2.07

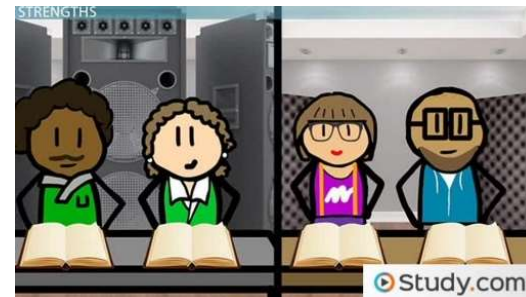
Average      1.832      2.068

# Experimental research: Between Subjects

**Independent Two Groups of test users:**

Each group uses only 1 of the systems

Condition 1	Condition 2
	



**e.g. A/B testing**







**Independent sample t-test**

# Experimental research: Matched Pairs

---

## Matched Pairs:

Matching each participant with someone who is similar to them, and placing them in different conditions.

Condition 1	Condition 2
  	  

e.g. Old version vs New version of a Web site



**Dependent sample t-test**

# Experimental research: *Within Subjects*

---

One group of test users

Each person uses both systems

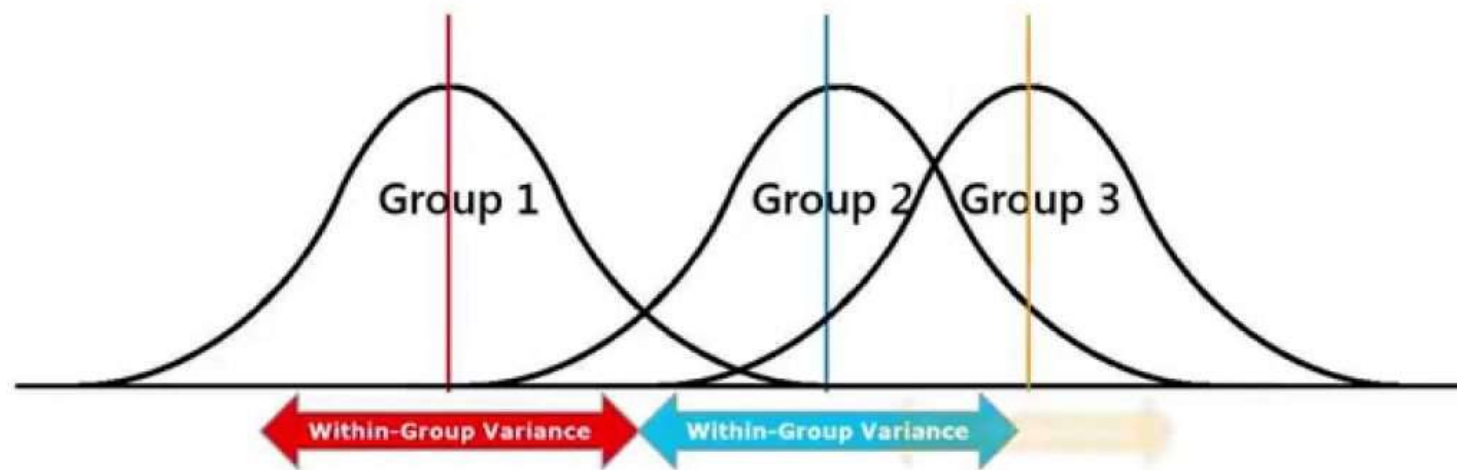
Condition 1	Condition 2
	

**Dependent sample t-test**

# Analysis of variance (“ANOVA”)

---

- the inferential statistic indicating the presence of a significant difference among the means of ***three or more groups***



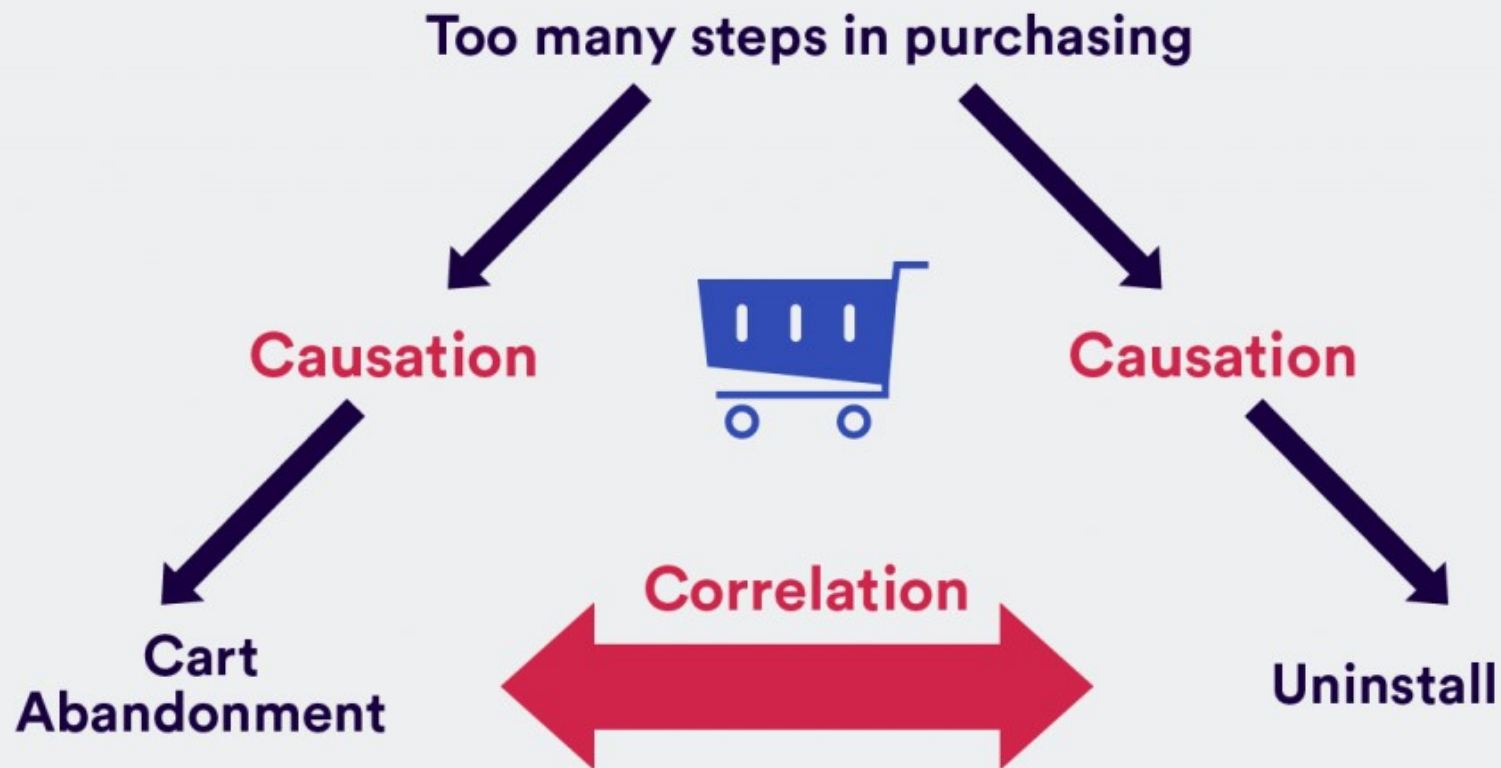
# Correlation

---

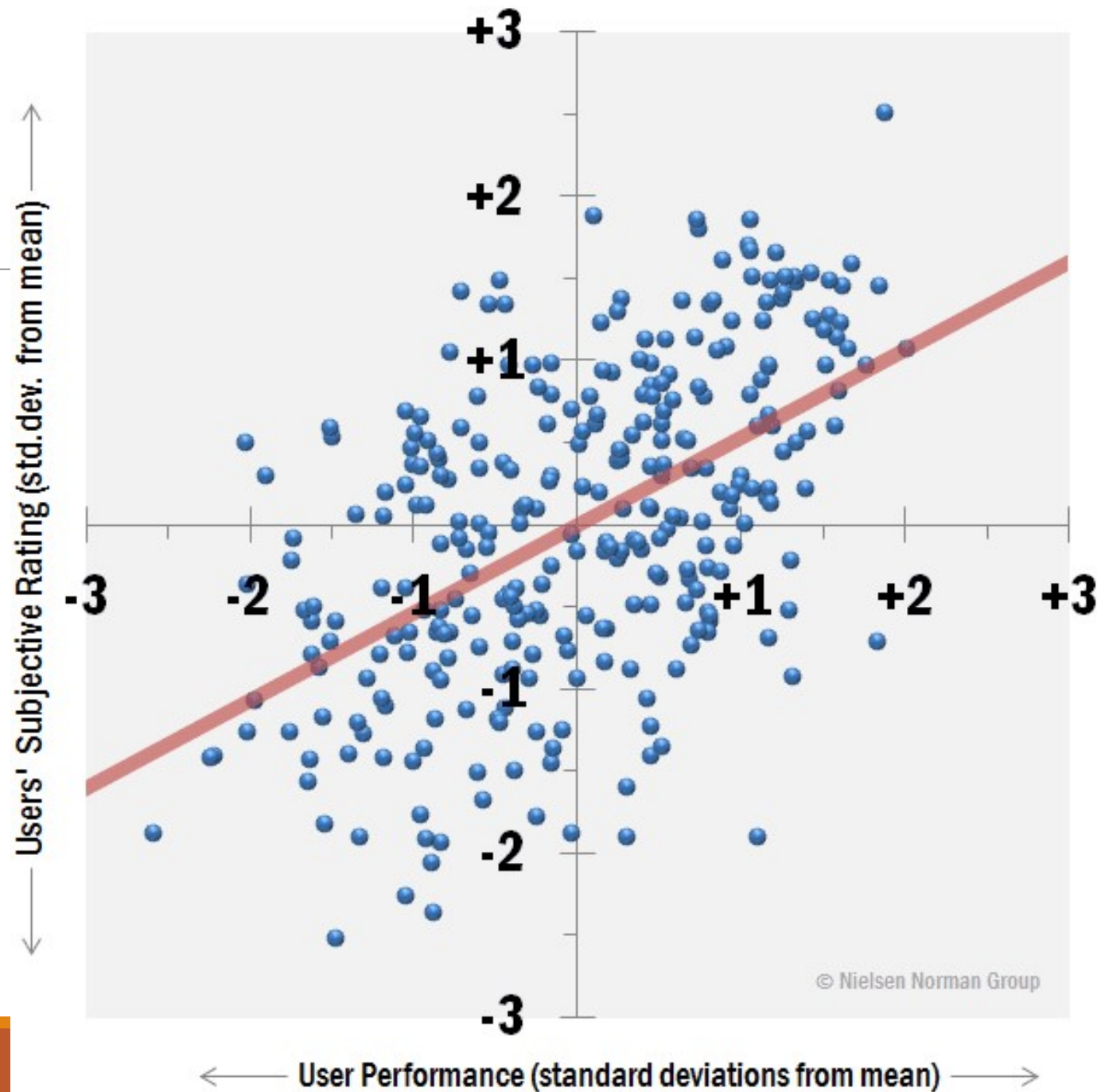
- The relationship between two variables of degree.
  - Positive: As one variable increases (or decreases) so does the other.
  - Negative: As one variable increases the other decreases.
- Magnitude or strength of relationship
  - -1.00 to +1.00
- Correlation does not equate to causation



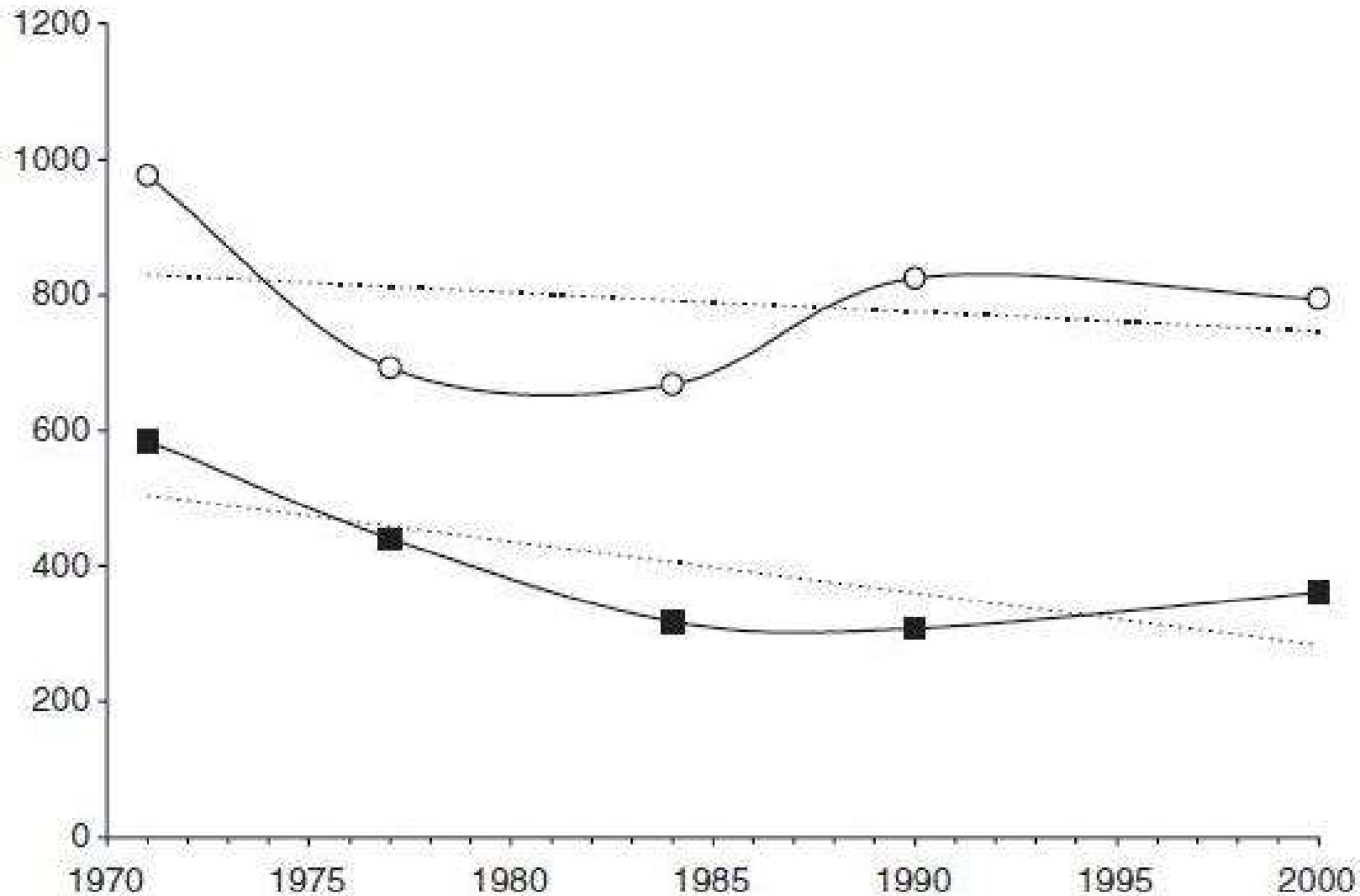
# Correlation Studies



# Correlation Studies



# Correlation isn't Causation



# Correlation isn't Causation

Example: Storks and birthrate

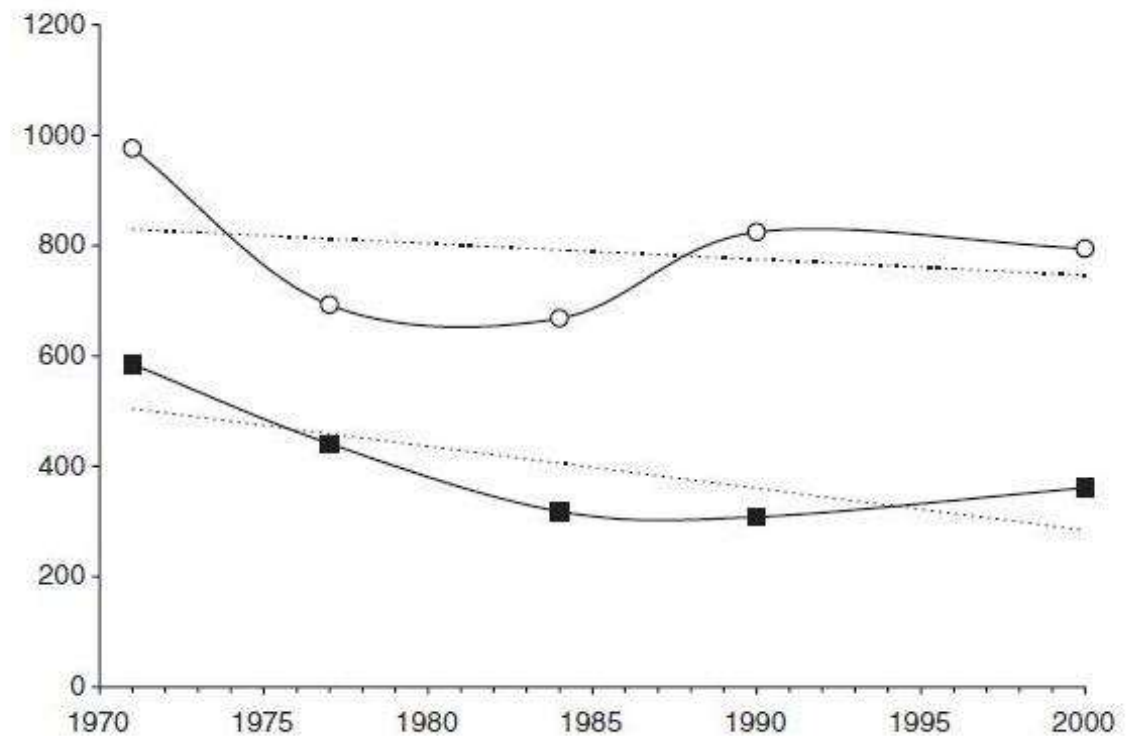
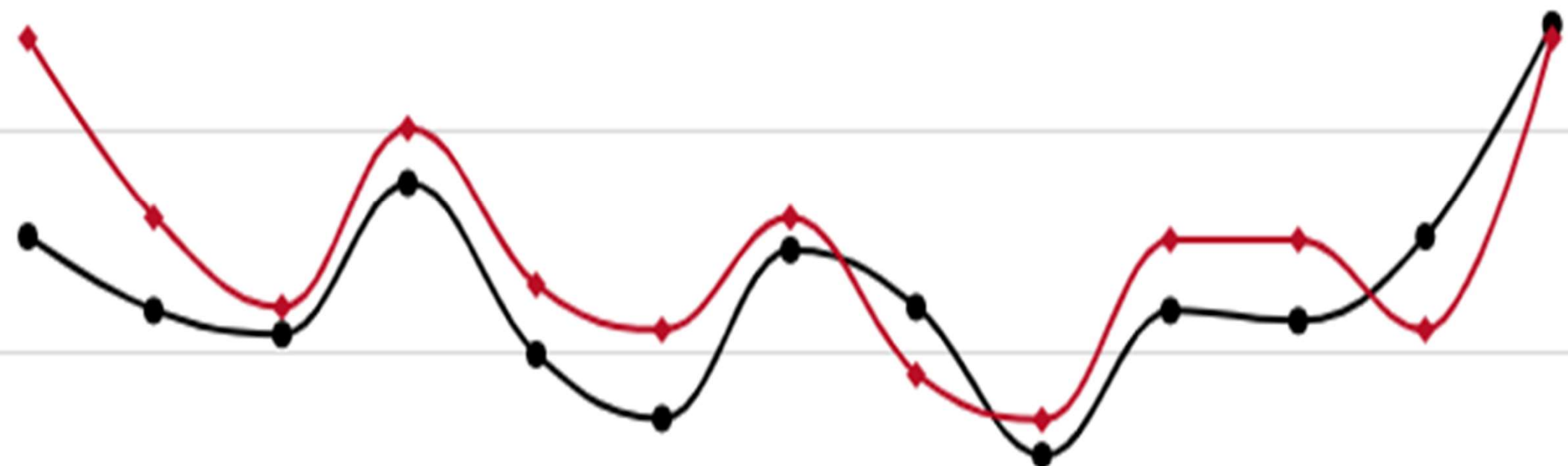


Figure 1. Storks and the birth rate in Lower Saxony, Germany (1971–2000). Open circles show yearly birthrates in hundreds in Lower Saxony. Full squares show numbers pairs of storks in Lower Saxony. Dotted lines represent linear regression trend ( $y = mx + b$ ).

Matthews, R. (2000), Storks Deliver Babies ( $p = 0.008$ ). Teaching Statistics, 22: 36-38. doi:[10.1111/1467-9639.00013](https://doi.org/10.1111/1467-9639.00013)

Correlation: 78.92% ( $r=0.78915$ )

1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009



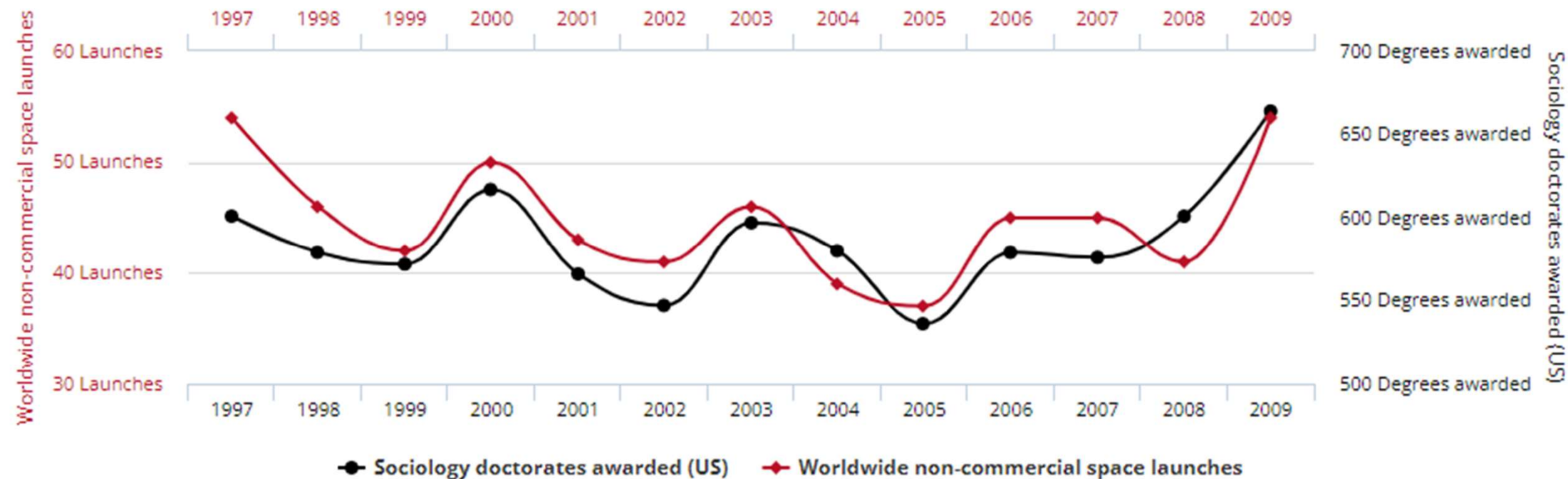
1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009

## Worldwide non-commercial space launches

correlates with

## Sociology doctorates awarded (US)

Correlation: 78.92% ( $r=0.78915$ )



# **Internal Validity:** accurately measure the relationship between the variables ?

---

- Identification, documentation, and elimination of confounds
- High, when there are no alternative explanations for your results
  - The variation of your dependent variable is caused by the variation of your independent variable
- Low, when there are experimental effects can be explained through confounds, bias, history effects, maturation, etc
  - The variation of your dependent variable can be explained by the variation of confounds
- We aim for high internal validity
- e.g. Usability testing of Akbank vs Vakıfbank with Sabancı students?

# External Validity

---

- The extent to which results can be generalized
- High, when results of the study can be transferred to the real world
  - e.g. does the sample represent the general population?
- Low when the results cannot be applied to the population or real-life situations outside of the research setting
  - ecological validity



# Final Report - structure

---

- Empirical
  - Gather data on an HCI topic
- Title

Abstract

Introduction

Method

Results

Discussion/Conclusion

References

# Results

---

- How have the data been treated?
- Text and graphs
- Statistics - descriptive/inferential
- Summarize the results

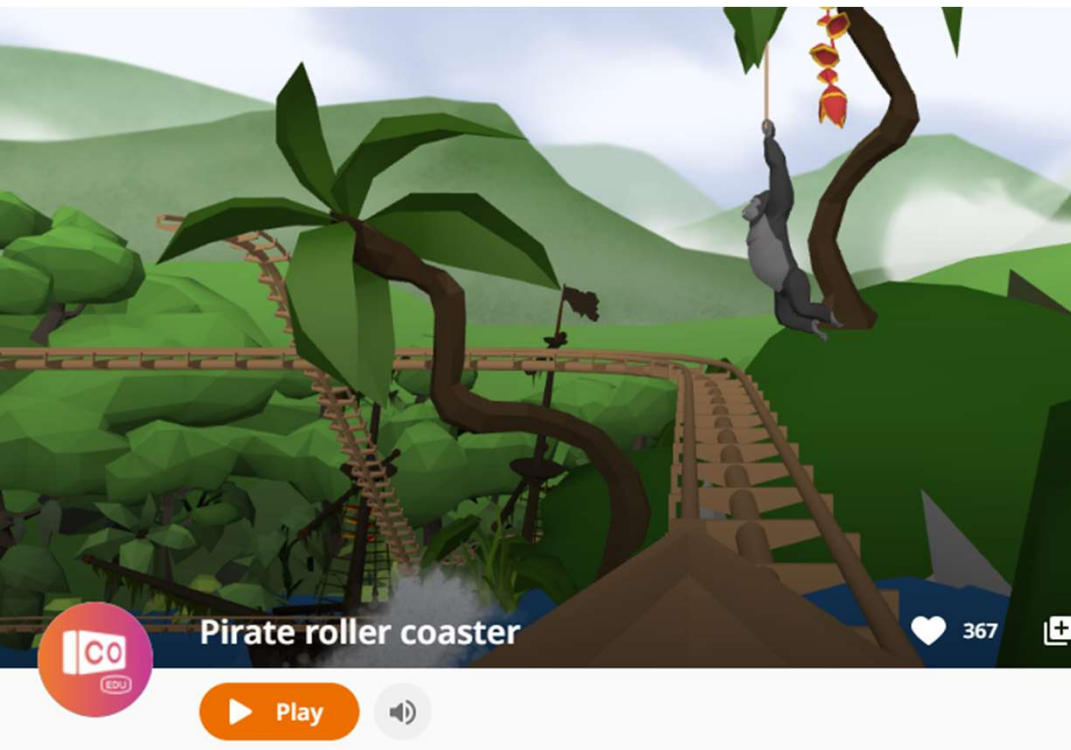
# Warnings

---

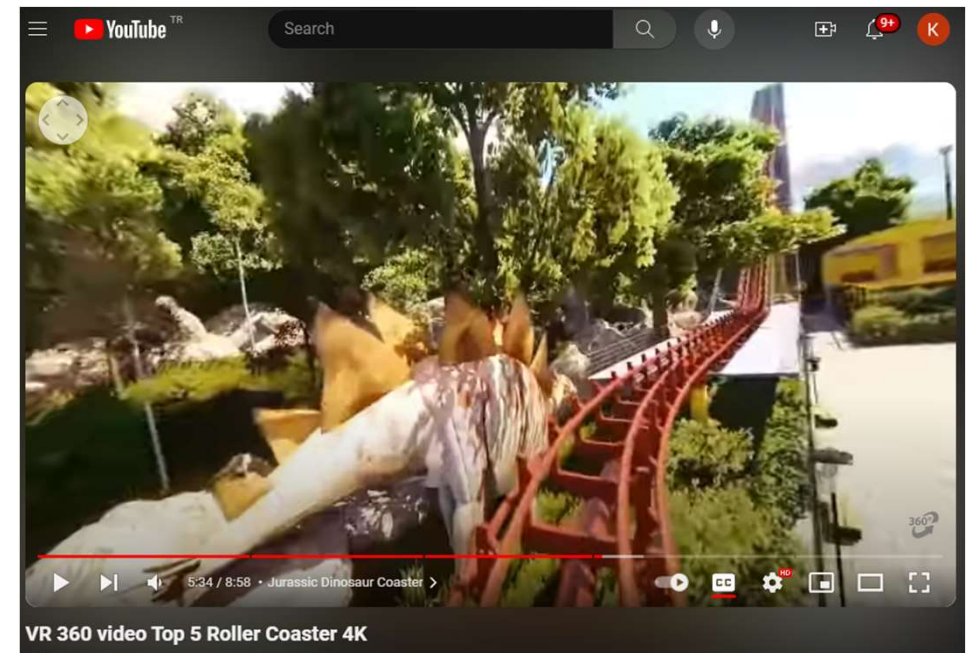
- Make a pilot test
- Keep the variables under control
  - E.g conduct tests in the same environment, use standard hardware, use standard instructions
- Participant characteristics / Persona
- Gender / Education / Background balance
- Task shuffling

# Assignment: VR Usability

- Group Assignment: **with your term project group, Due December 26th**
- 



<https://edu.cospaces.io/VMX-SBG>



<https://youtu.be/hNAbQYU0wpg?si=-Pwn4fKldEnDBku9&t=330>

<https://youtu.be/kKQgFFz66a4?feature=shared&t=15>

- **1-Try both applications within your group first. Be familiar with application and the test procedure**
- **2- Users:** Prepare one persona for your target user group. Find **minimum 4 users** to test both applications. Keep gender and background balance. Make sure they have **no** previous VR experience. Before the test collect demographics data from each user.

---

- **3- Task:** Ask them to watch both VR applications. Task order is important, if the first user watches Application-1 first, the second user has to watch Application-2 first.
- **4- Context:** Conduct the test while the end-users sitting in a quiet place
- **5- Tool:** Conduct the test with the same mobile phone for all users.
- **6- During the test (Quantitative and Qualitative data):** Record reactions of users during the test.
- **7- Post interview (Qualitative data):** Prepare interview questions to compare and assess the level of users' preferences in the two Virtual Reality (VR) roller coaster applications. Evaluate factors contributing to immersion and identify strengths and weaknesses in creating a captivating VR experience.
- **8- Post-test (Quantitative data):** Ask subjects to complete SUS (System Usability Scale) for both applications (One for YouTube and one for CoSpaces). Use English version of it. (It is on SuCourse, Week-10)



#### Ferhat Hakkında Özet Bilgi

- Yönetici
- Orta seviye bilgisayar kullanıcısı
- E-posta, web, çevrim-ichi bankacılık kullanır
- Laptop kullanıcısı
- Evde ve iş yerinde kullanıyor

**İşletme:** Dilşat Balıkçılık Ltd.

**Kişi:** Ferhat Çağıltay

**Pozisyon:** Şirket sahibi

Ferhat 43 yaşında, evli ve bir çocukludur. Ziraat Fakültesi, Su ürünleri bölümü mezunu olup, halen bu alanda kurduğu firmada balıkçılık yapmaktadır. Bilgisayarı yazı yazmak, haberleşmek ve şirketinin banka hesaplarını takip etmek için kullanmaktadır. Genellikle günde 1 ila 1.5 saat bilgisayar başında harcamakta ve bu zamanın yarısını İnternet'te geçirmektedir.

Hem evde ve hem de iş yerinde İnternet bağlantısı ADSL hattı üzerinden olup, dizüstü bilgisayar kullanmaktadır.

Ferhat kendisini, iş için bilgisayar kullanan, orta seviye bilgisayar kullanıcısı olarak belirtmektedir.

Çevrimiçi bankacılık servislerinin ticari olanlarını gün aşırı kullanmakta, hesap kontrolü, yatırım hesabı ve EFT işlemlerini sık sık yapmaktadır.

# Lets try the VR

---