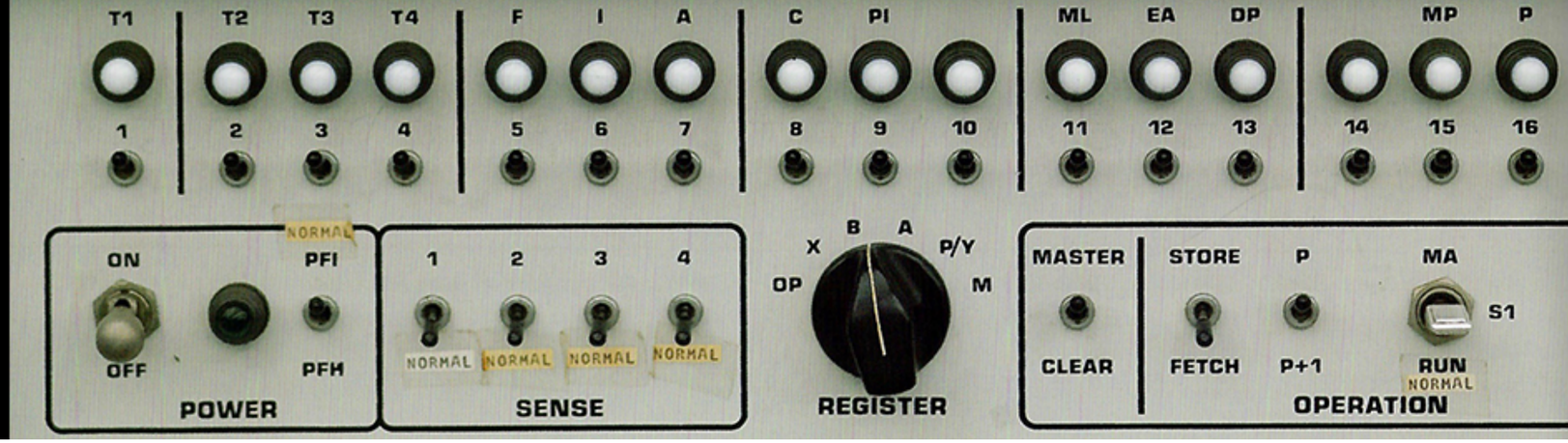
CS160

USER INTERFACE DESIGN

FALL 2018



DATA ANALYSIS

10 OCT 2018

ERIC PAULOS
www.paulos.net

UNIVERSITY OF CALIFORNIA Berkeley

MIDTERM ON 15 OCT

In class – Actually in Sibley Auditorium

Watch Piazza for details

80 minutes

Closed book & notes

If you are registered with the DSP office, you should have received email from us about exam accommodations. All accommodations finalized today please.

MIDTERM

HKN has previous midterm for practice

come early so we can start on time

all you need is a pen or pencil to write with

The midterm will cover all aspects of the course through Wednesday's lecture. This includes, readings, lectures, assignments, section, etc. We may ask design questions, code questions, etc.

There will be a midterm Review in sections this Friday

If you find a question ambiguous, document the ambiguity. Indicate the way you interpreted the question in a set of separate sentences next to the question. The questions on the exam are not intended to be ambiguous, but sometimes another meaning is interpreted by the examinee that we did not take into consideration.

If you are registered with the DSP office and have special needs, you should have received email from me about exam accommodations via bCourses. You must contact me by TODAY if there is a problem in any accommodations details or accommodations made (or not made) to you.

MIDTERM ATTENDANCE

I know this won't happen but I'm putting it here so it is clear there is policy:

You must attend the midterm

There is not a makeup midterm exam date

If you have any reason to believe you may miss the midterm (i.e. you have a court appearance, you have difficult travel plans, you have a planned doctors appointment before class that may run over, you have a job interview, you have been called to testify before Congress, etc), you must let me know by class Wed 10 Oct. **We will not grant excuses for issues that come up after Wednesday** and you will be given a zero for the exam if you do not attend.

I will not read or respond to any requests concerning issues of why you cannot attend or will not be able to attend the midterm after end of class Wed 10 Oct.

MIDTERM ATTENDANCE

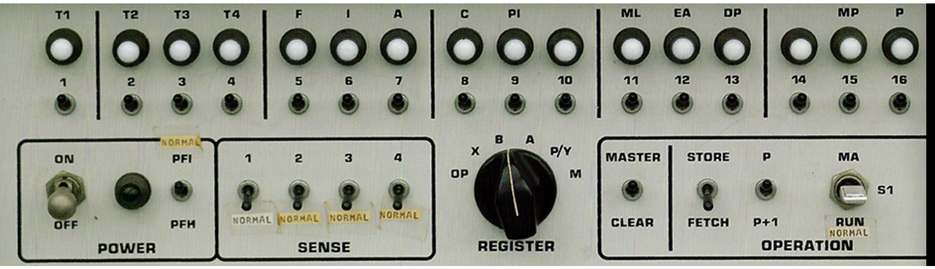
If you have a health or medical emergency and unable to make the midterm or decide not to come:

- You must tell us immediately via Piazza post to all instructors
- You must meet with me (Professor Paulos) as soon as possible to discuss your circumstances
- You will be given, at the complete discretion of the instructor, either a zero for the midterm, an oral exam of up to 3 hours (on a date set by the instructor that is not negotiable), or a 3 hour final exam during our scheduled final exam time covering the material from the entire semester. The grade on this exam will take the place of your midterm.
- Legitimate health related emergencies (as determined by the instructor) will not be given a zero but will be subject to the other two options at the discretion of the instructor.

MIDTERM ATTENDANCE

Should I come to the midterm (cheat sheet):

- I’m exhausted and need to sleep — ATTEND
- I think I’m starting to get a cold — ATTEND
- I am profusely sick and vomiting all over my bed — Goto Tang / DO NOT ATTEND
- I broke my leg — Call 911 / DO NOT ATTEND
- I am bleeding profusely. Help! — Call 911 / DO NOT ATTEND
- I’m not sure, I’ll email Professor Paulos — ATTEND as I will not be able to respond to any emails that arrive concerning the midterm attendance after 10 Oct



MANAGING STUDY PARTICIPANTS

RUN THE EXPERIMENT

Always pilot it first!

Reveals unexpected problems

Can’t change experiment design after starting it

Always follow same steps – use a checklist

Get consent from subjects

Debrief subjects afterwards

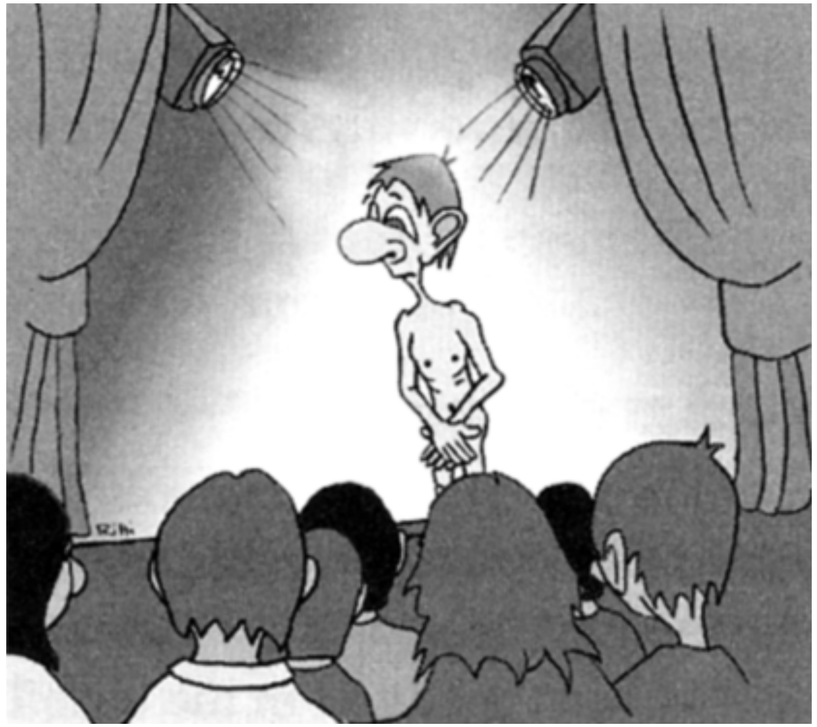
THE PARTICIPANTS' STANDPOINT

Testing is a distressing experience

Pressure to perform

Feeling of inadequacy

Looking like a fool in front of your peers, your boss, ...



(from "Paper Prototyping" by Snyder)

THE THREE BELMONT PRINCIPLES

Respect for Persons

Have a meaningful consent process: give information, and let prospective subjects freely chose to participate

Beneficence

Minimize the risk of harm to subjects, maximize potential benefits

Justice

Use fair procedures to select subjects

Burdens and benefits shared equitably
(balance burdens & benefits)

To ensure adherence to principles, most schools require Institutional Review Board approval of research involving human subjects.

THE THREE BELMONT PRINCIPLES

Respect for persons

protecting the autonomy of all people and treating them with courtesy and respect and allowing for informed consent. Researchers must be truthful and conduct no deception

Beneficence

The philosophy of “Do no harm” while maximizing benefits for the research project and minimizing risks to the research subjects

Justice

ensuring reasonable, non-exploitative, and well- considered procedures are administered fairly — the fair distribution of costs and benefits to potential research participants — and equally.

RESPECT FOR PERSONS

Treat individuals as autonomous agents

Persons with diminished autonomy are entitled to protection

Applications

Participation should be voluntary

Participants should be fully informed of the costs and benefits of participation

BENEFICENCE

Do not harm

Maximize the possible benefits and minimize the possible harms

Applications

Systematic analysis of the risks and benefits of the research to both the individual and to society at large

JUSTICE

Who should bear the burdens of research and who should receive the benefits?

To each person an equal share

To each person according to individual need

To each person according to individual effort

To each person according to societal contribution

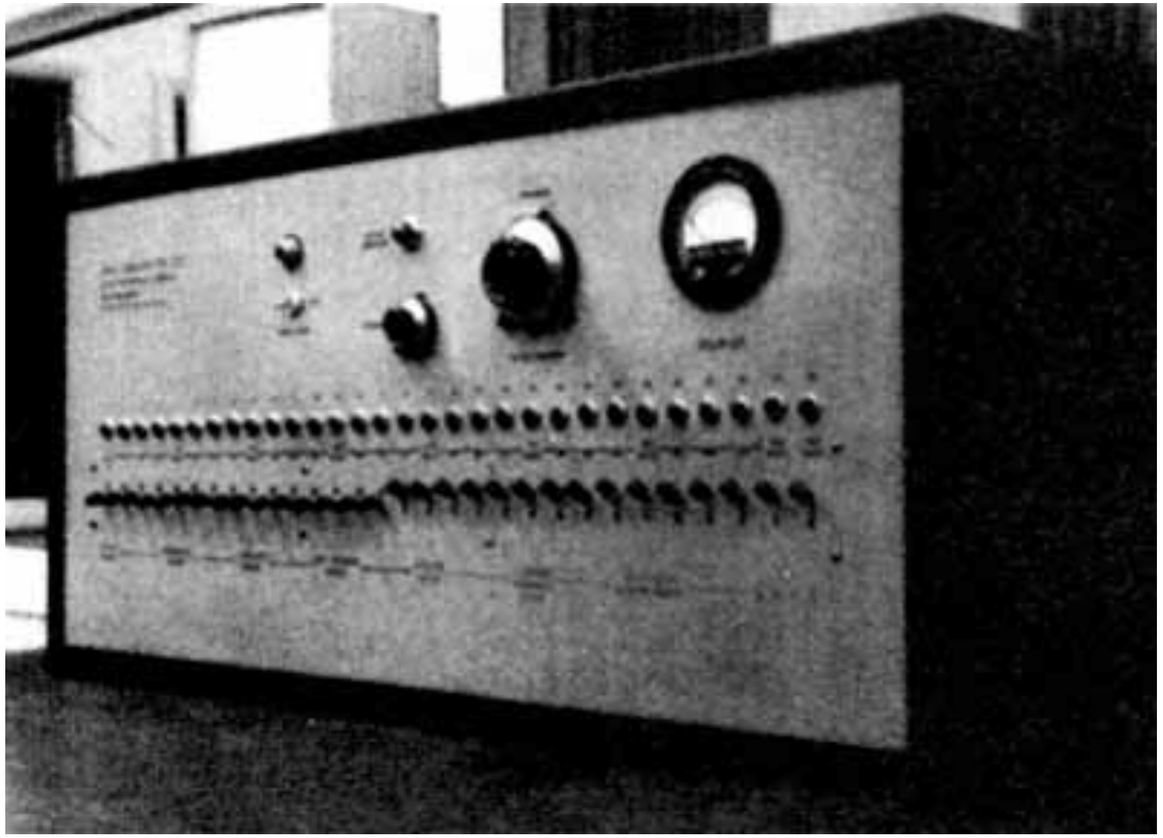
To each person according to merit

Application

Selection of research participants

MILGRAM OBEDIENCE TO AUTHORITY

1961 Experiment by Stanley Milgram



ETHICS: STANFORD PRISON EXPERIMENT

1971 Experiment by Phil Zimbardo at Stanford

24 Participants – half prisoners, half guards ($15 a day)

Basement of Stanford Psychology building turned into mock prison

Guards given batons, military style uniform, mirror glasses,…

Prisoners wore smocks (no underwear), thong sandals, pantyhose caps

ETHICS: STANFORD PRISON EXPERIMENT

Experiment quickly got out of hand

Prisoners suffered and accepted sadistic treatment

Prison became unsanitary/inhospitable

Prisoner riot put down with use of fire extinguishers

Guards volunteered to work extra hours

Zimbardo terminated experiment early

Grad student Christina Maslach objected to experiment

Important to check protocol with ethics review boards

Christina MASLACH

Professor
Chair of the Academic Senate
Contact Information
maslach@berkeley.edu
Office: 3325 Tolman Hall
Office Hours: by appointment
http://psychology.berkeley.edu/maslach%40socialpsychology.org
Ph.D., Stanford University
Curriculum Vitae: CV 2012.pdf
Research Interests: Social: job burnout and health psychology; individuation and dissent, gender roles

Research Areas: Social-Personality

ETHICS

Was it useful?

"...that's the most valuable kind of information that you can have and that certainly a society needs it" (Zimbardo)

Was it ethical?

Could we have gathered this knowledge by other means?

ETHICS (MORE RECENTLY)

In 2001, a faculty member from the business school of a major university designed a study to see how restaurants would respond to complaints from putative customers. As part of the project, the researcher sent letters to restaurants falsely claiming that he and/or his wife had suffered food poisoning that ruined their anniversary celebration. The letters disclaimed any intention of contacting regulatory agencies and stated that the only intent was to convey to the owner what had occurred “in anticipation that you will respond accordingly.” Restaurant owners were understandably upset and some employees lost their jobs before it was revealed that the letter was a hoax.

ETHICS (EVEN MORE RECENTLY)

The Study

All Facebook users who spoke English qualified

Two groups: positive and negative emotions

Positive/negative posts where then suppressed from the news feed

689,003 participants randomly selected by user id

Saw an impact

When **positive posts withheld** the **participant's posts got more negative**

When **negative posts withheld** the **participants posts got more positive**

Withdrawal effect: people who saw less emotion posts less likely to express themselves for several days

ETHICS (EVEN MORE RECENTLY)

In June 2014 researchers from Facebook altered the news feed algorithm for 689,003 users to skew the presence of positive or negative posts. They then tracked subsequent posts from those users by using positive or negative keywords.

"In addition to helping people see and find things that you do and share, we may use the information we receive about you...for internal operations, including troubleshooting, data analysis, testing, research and service improvement."

Institutions that receive federal funding are required to abide by a federal policy called the "Common Rule," which protects human experiment subjects by ensuring that they know about the study and that they understand the risks involved. It also requires institutional review boards at universities and hospitals to approve the way subjects of biomedical or behavioral studies are treated.

ETHICS (EVEN MORE RECENTLY)

Lead researcher and Facebook data scientist Adam Kramer took to Facebook to defend the study:

“We felt that it was important to investigate the common worry that seeing friends post positive content leads to people feeling negative or left out. At the same time, we were concerned that exposure to friends' negativity might lead people to avoid visiting Facebook,” Kramer wrote.

He went on to explain that the “actual impact on people” was the minimal needed to conclude that Facebook feeds influenced users’ emotions. Though they expected happy news would make people feel sad, they found that people with a little more positive news in their feeds included more happy words in their posts.

“Having written and designed this experiment myself, I can tell you that our goal was never to upset anyone,” he wrote in the post. “I can understand why some people have concerns about it, and my coauthors and I are very sorry for the way the paper described the research and any anxiety it caused. In hindsight, the research benefits of the paper may not have justified all of this anxiety.”

BENEFICENCE: EXAMPLE

MERL DiamondTouch

User capacitively coupled to table through seating pad

No danger for normal users, **but possibly increased risk for participants with pacemakers**

Inform subjects in consent!

PRIVACY AND CONFIDENTIALITY

Privacy — having control over the extent, timing, and circumstances of sharing oneself with others.

Confidentiality — the treatment of information that an individual has disclosed with the expectation that it will not be divulged

Examples where privacy could be violated or confidentiality may be breached in HCI studies?

TREATING SUBJECTS WITH RESPECT

Follow human subject protocols

Individual test results will be kept confidential

Users can stop the test at any time

Users are aware (and understand) the monitoring technique(s)

Their performance will not have implications on their life

Records will be made anonymous

Use standard informed consent form

Especially for quantitative tests

Be aware of legal requirements

CONDUCTING THE EXPERIMENT

Before the experiment

Have them read and sign the consent form

Explain the goal of the experiment in a way accessible to users

Be careful about the demand characteristic
(Participants biased towards experimenter's hypothesis)

Answer questions

During the experiment

Stay neutral

Never indicate displeasure with users performance

After the experiment

Debrief users (Inform users about the goal of the experiment)

Answer any questions they have

MANAGING SUBJECTS

Don’t waste users’ time

Use pilot tests to debug experiments, questionnaires, etc…

Have everything ready before users show up

Make users comfortable

Keep a relaxed atmosphere

Allow for breaks

Pace tasks correctly

Stop the test if it becomes too unpleasant

IF YOU WANT TO LEARN MORE...

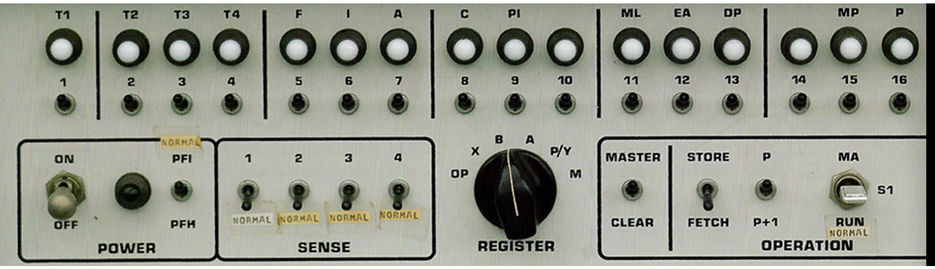
Online human subjects certification courses:

E.g., http://phrp.nihtraining.com/users/login.php

The Belmont Report: Ethical Principles and Guidelines for the protection of human subjects of research

1979 Government report that describes the basic ethical principles that should underly the conduct of research involving human subjects

http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html



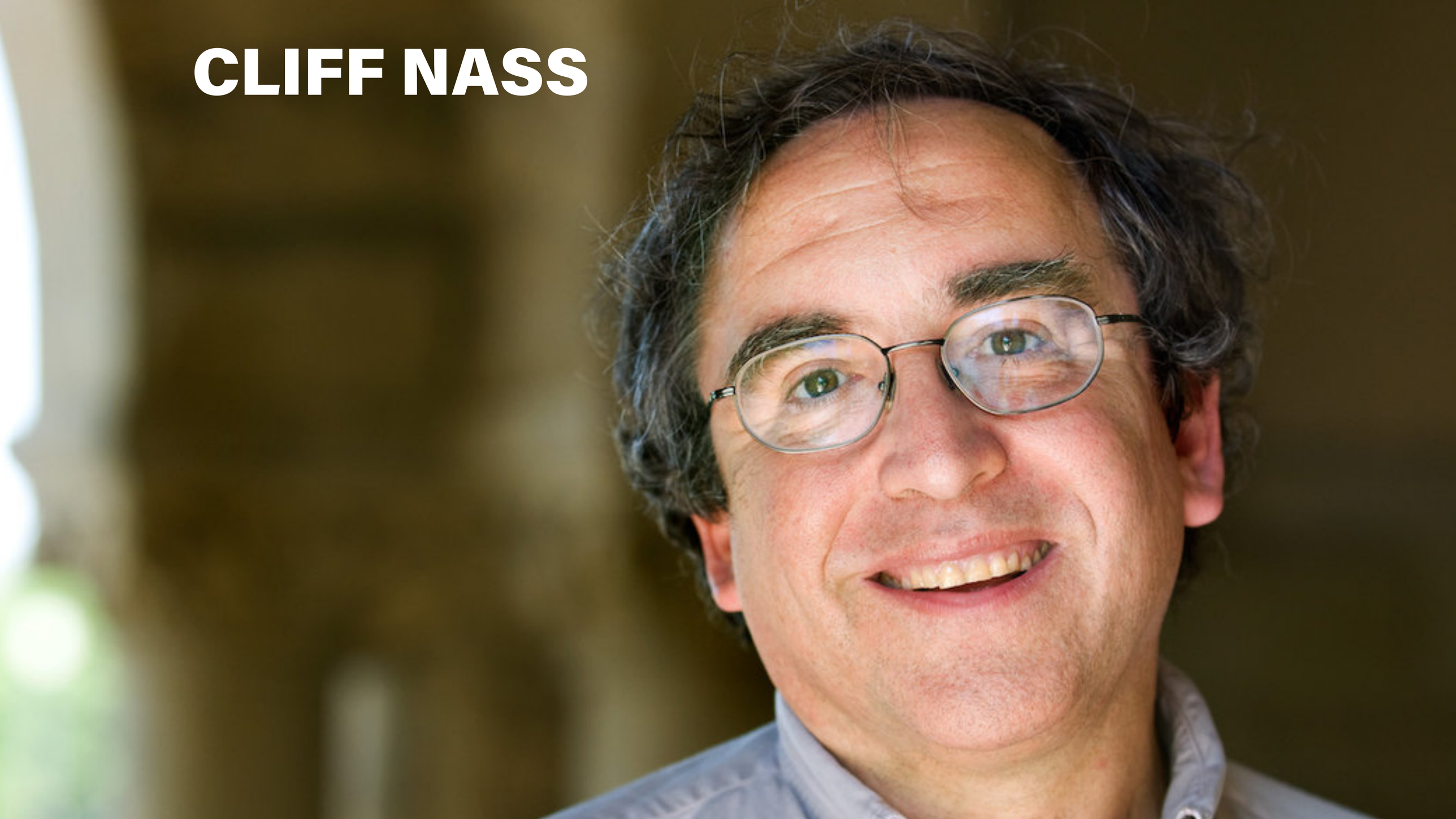
MULTI-TASKING

MULTI-TASKING

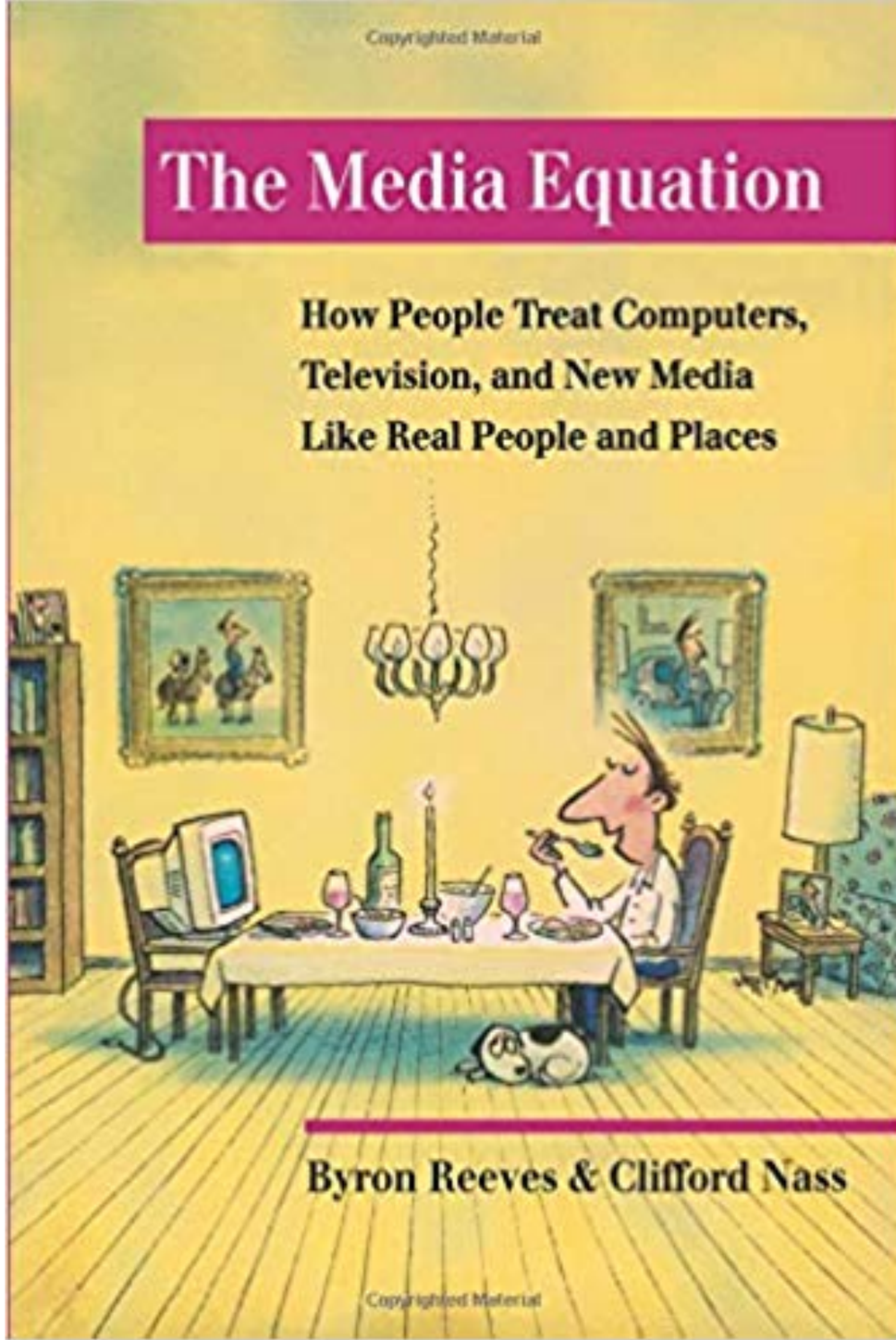
Social scientists have long assumed that it’s impossible to process more than one string of information at a time.

The brain just can’t do it.

But many researchers have guessed that people who appear to multitask must have superb control over what they think about and what they pay attention to.


CLIFF NASS

The Media Equation is a general communication theory that claims that people tend to treat computers and other media as if they were either real people or real places



ALONE TOGETHER

MULTI-TASKING

HCI Researchers split their subjects into two groups:

1) those who regularly do a lot of media multitasking

2) those who don’t

MULTI-TASKING

In one experiment, the groups were shown sets of two red rectangles alone or surrounded by two, four or six blue rectangles. Each configuration was flashed twice, and the participants had to determine whether the two red rectangles in the second frame were in a different position than in the first frame.

MULTI-TASKING

They were told to ignore the blue rectangles, and the **low multitaskers had no problem doing that**. But the **high multitaskers were constantly distracted by the irrelevant blue images.** Their performance was horrible.

MULTI-TASKING

Because the high multitaskers showed they couldn’t ignore things, the researchers figured they were better at storing and organizing information. **Maybe they had better memories**.

MULTI-TASKING

The second test proved that theory wrong.

After being shown sequences of alphabetical letters, the high multitaskers did a lousy job at remembering when a letter was making a repeat appearance.

...The low multitaskers did great! The high multitaskers were doing worse and worse the further they went along because they kept seeing more letters and had difficulty keeping them sorted in their brains.

MULTI-TASKING

If the heavy multitaskers couldn’t filter out irrelevant information or organize their memories, perhaps they excelled at switching from one thing to another faster and better than anyone else.

MULTI-TASKING

Wrong again, the study found.

The test subjects were shown images of letters and numbers at the same time and instructed what to focus on. When they were told to pay attention to numbers, they had to determine if the digits were even or odd. When told to concentrate on letters, they had to say whether they were vowels or consonants.

Again, the heavy multitaskers underperformed the light multitaskers.

“They couldn’t help thinking about the task they weren’t doing”

“The high multitaskers are always drawing from all the information in front of them. They can’t keep things separate in their minds.”

MULTI-TASKING

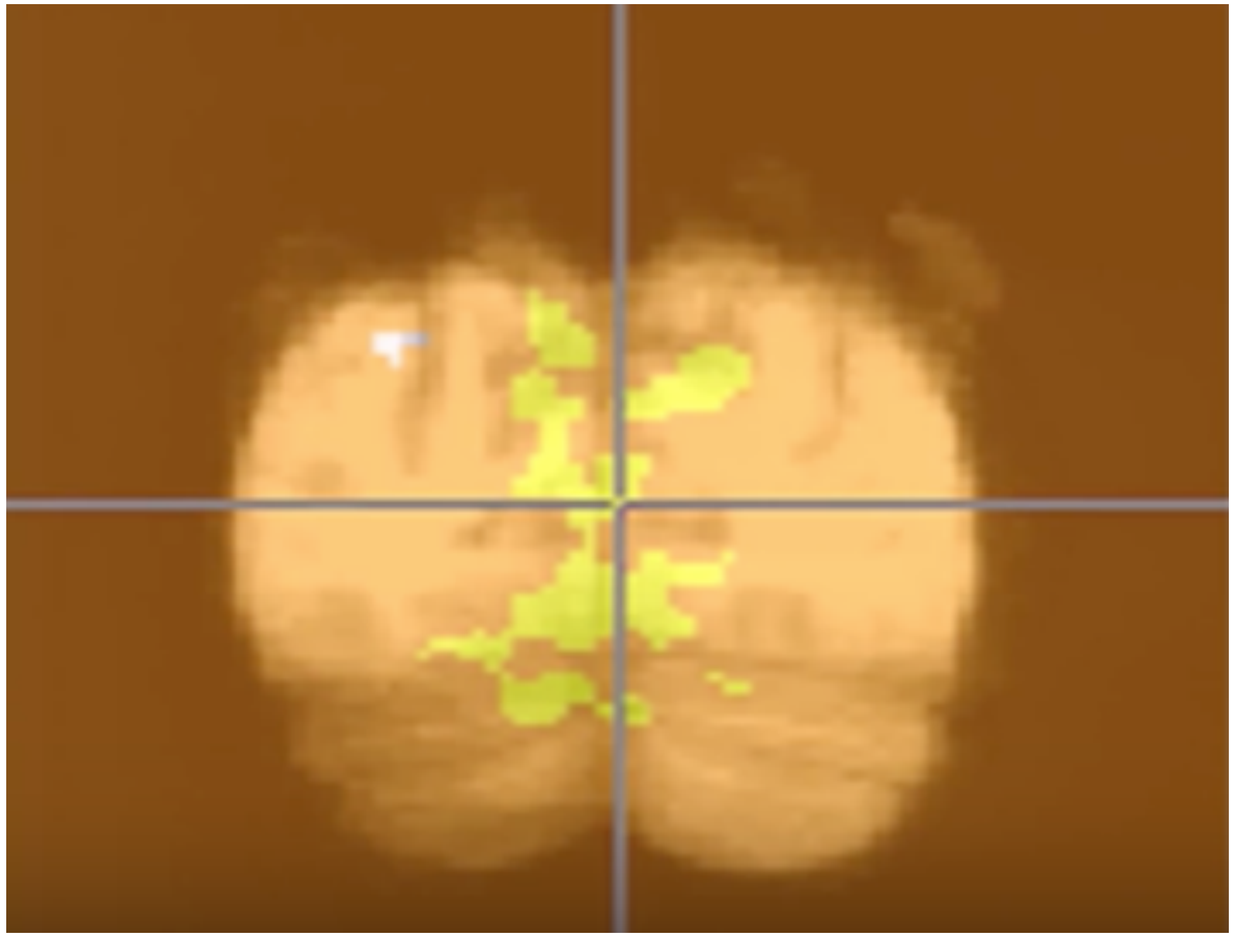
When multitaskers are in situations where there are multiple sources of information coming from the external world or emerging out of memory, they're not able to filter out what's not relevant to their current goal. That failure to filter means they're slowed down by that irrelevant information.

MULTI-TASKING

Poor filtering

Ineffective memory management

Suckers for irrelevancy



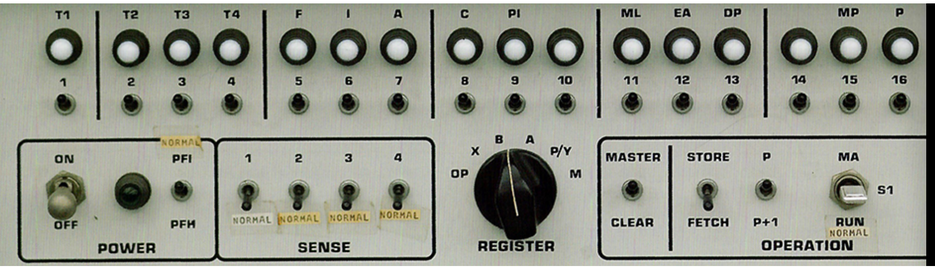
MULTI-TASKING

Frequent multitaskers

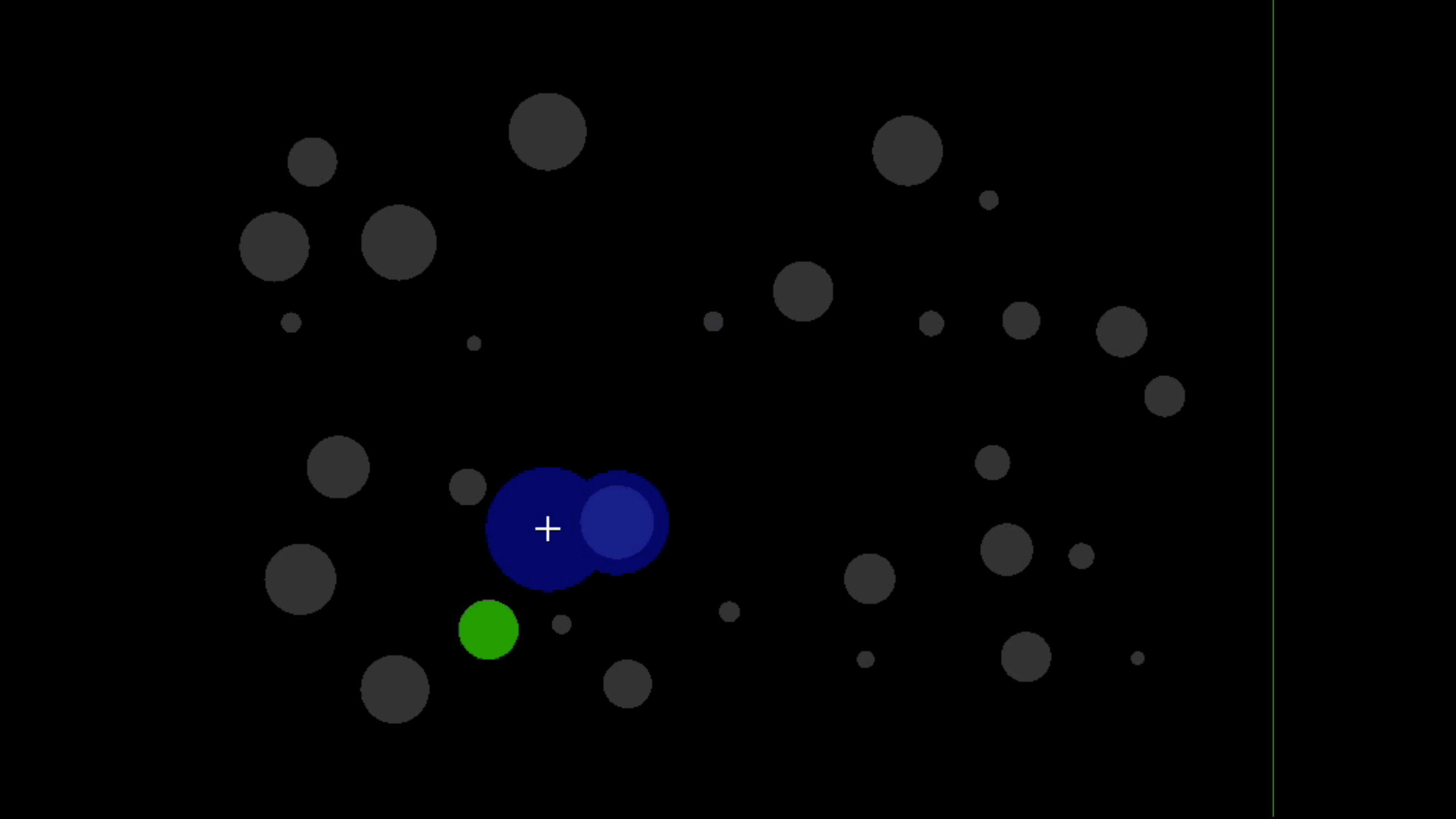
- Used media when face-to-face
- Feel less normal
- More bad influences (friends)
- Less Sleep

Face-to-Face Interaction

- Focused on other person
- Greater social success
- Felt more normal
- Had better friendships
- Got more sleep



DATA ANALYSIS



```
1362884269,normal,10,1,t5,1474
1362884269,normal,10,1,t6,979
1362884269,normal,10,1,t7,944
1362884269,normal,10,1,t8,966
1362884269,normal,10,1,t9,931
1362884269,normal,10,1,t10,926
1362884269,normal,10,1,t11,1047
1362884269,normal,10,1,t12,949
1362884269,normal,10,1,t13,980
1362884269,normal,10,1,t14,996
1362884269,normal,10,1,t15,930
1362884269,normal,10,1,t16,935
1362884269,normal,10,1,t17,797
1362884269,normal,10,1,t18,899
1362884269,normal,10,1,t19,945
1362884269,normal,10,1,t20,945
1362884302,bubble,10,0,t1,749
1362884302,bubble,10,0,t2,631
1362884302,bubble,10,0,t3,694
```

START BY COUNTING

```
4140 trials total


normal:
mean time 955.4 ms,
mean errors 1.486


bubble:
mean time 763.9 ms,
mean errors 0.402
```

START BY COUNTING — COMBINATIONS

```
54 users completed condition normal, size 10
mean time: 1113.25 ms, mean errors: 1.889
median time: 1067 ms, median errors: 1

51 users completed condition normal, size 30
mean time: 788.33 ms, mean errors: 1.059
median time: 754 ms, median errors: 1

52 users completed condition bubble, size 10
mean time: 809.96 ms, mean errors: 0.404
median time: 766 ms, median errors: 0

50 users completed condition bubble, size 30
mean time: 716.01 ms, mean errors: 0.020
median time: 692 ms, median errors: 0
```

DESCRIPTIVE STATISTICS

Continuous data

Central tendency

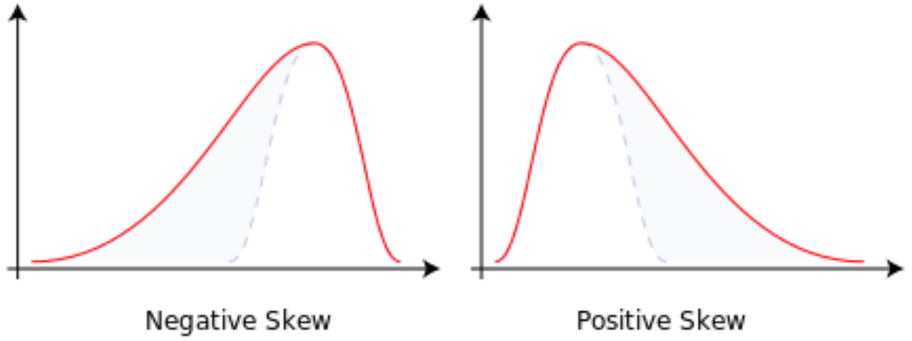
- mean, median, mode

Dispersion

- Range (max-min)
- Standard deviation

Shape of distribution

- Skew
- Kurtosis

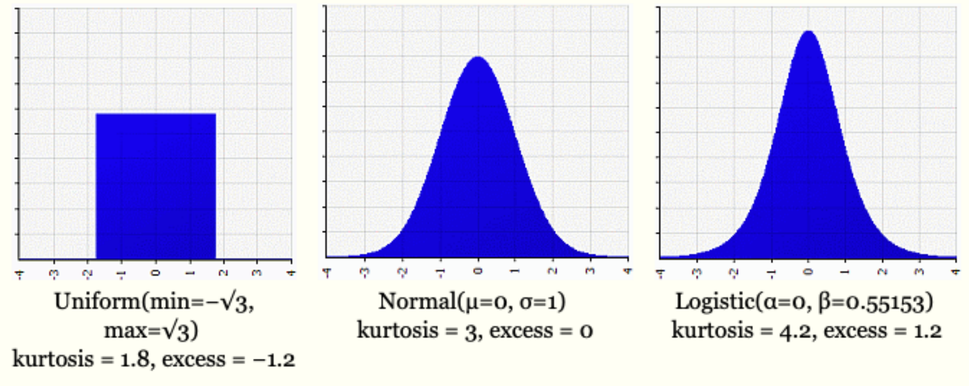


$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

Mean

Categorical data

Frequency distributions



$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

Standard Deviation

UNDERSTANDING YOUR DATA

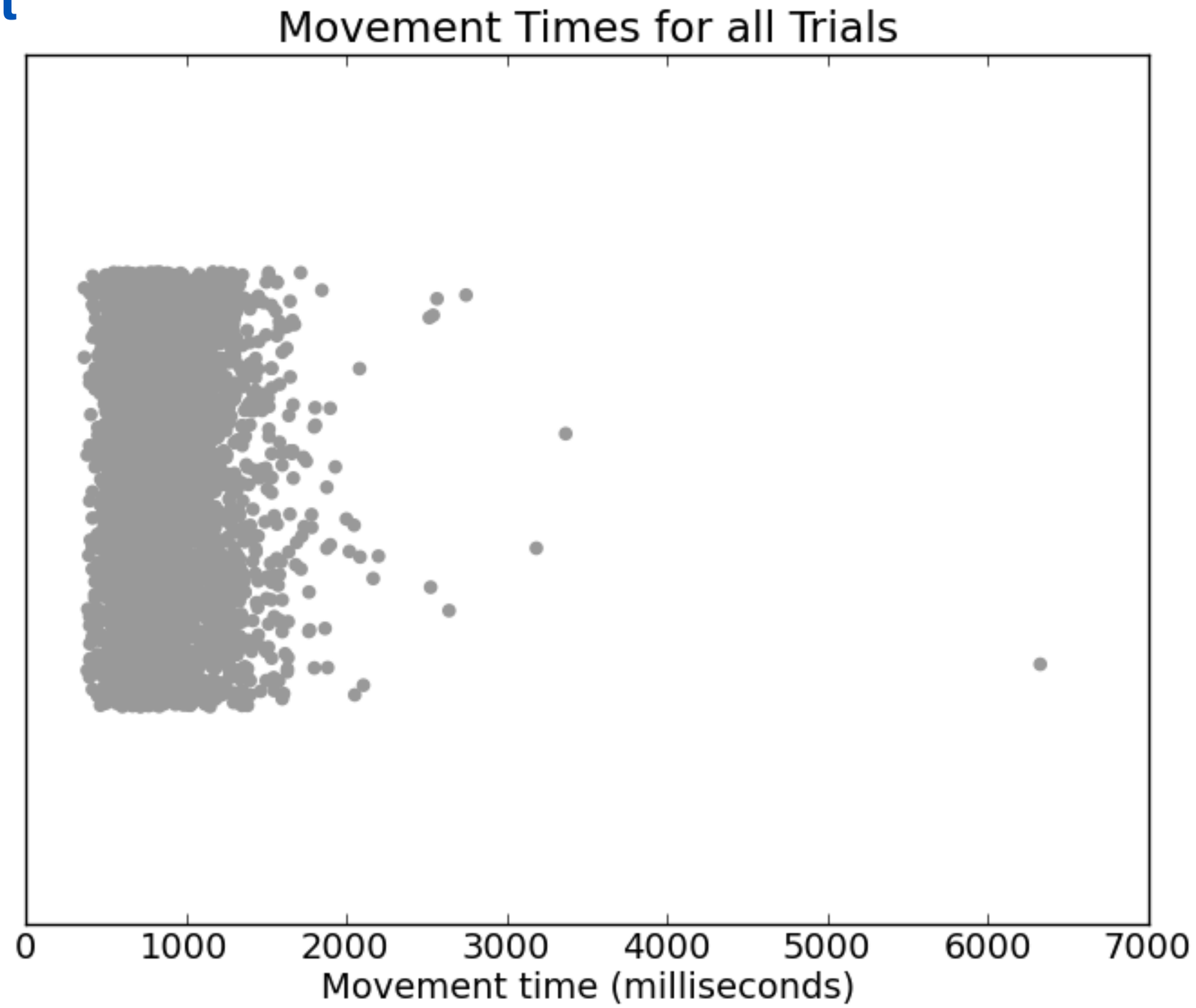
Exploratory Data Analysis

Look at your data from different perspectives to get better intuition for it.

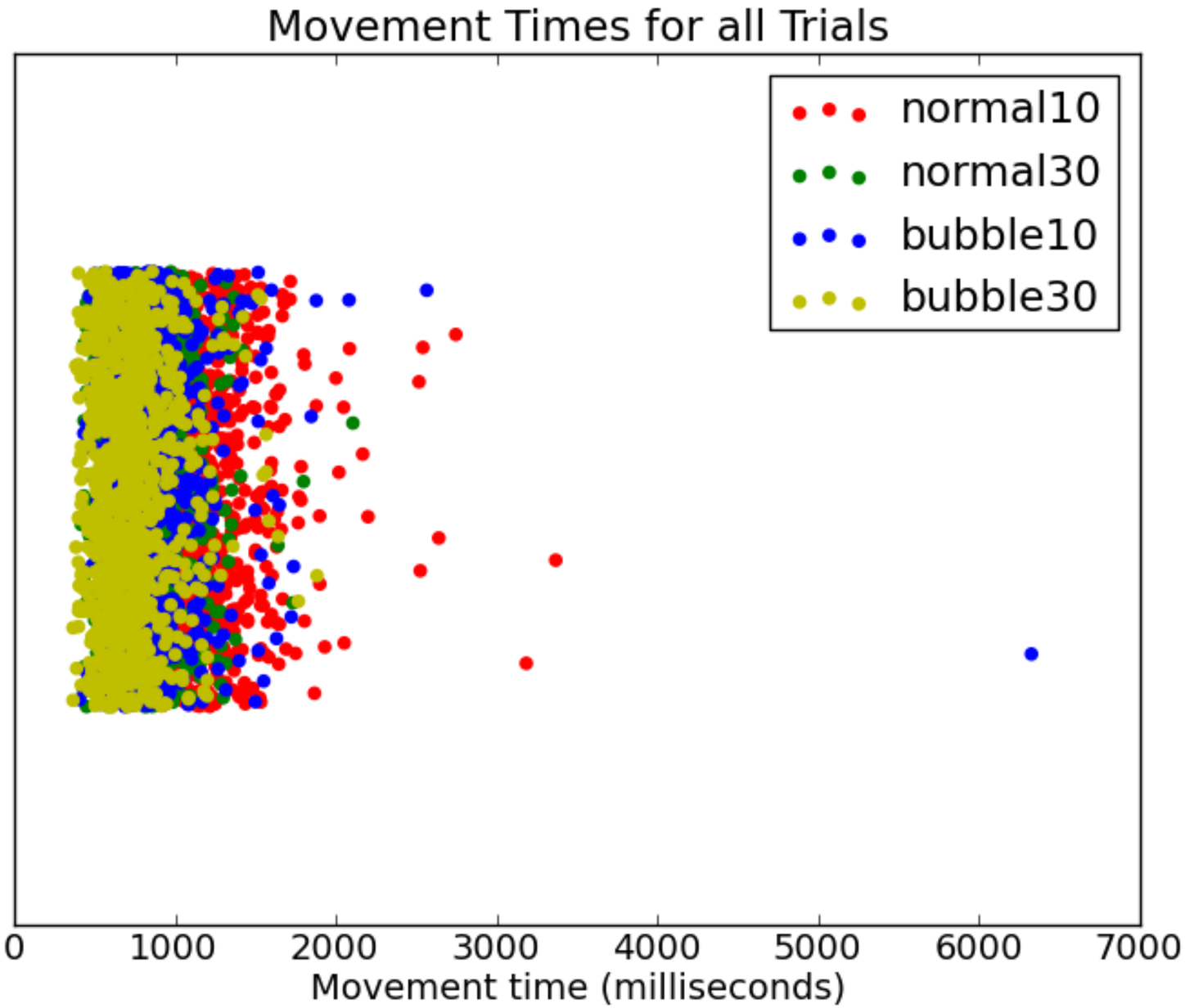
Show the raw data!

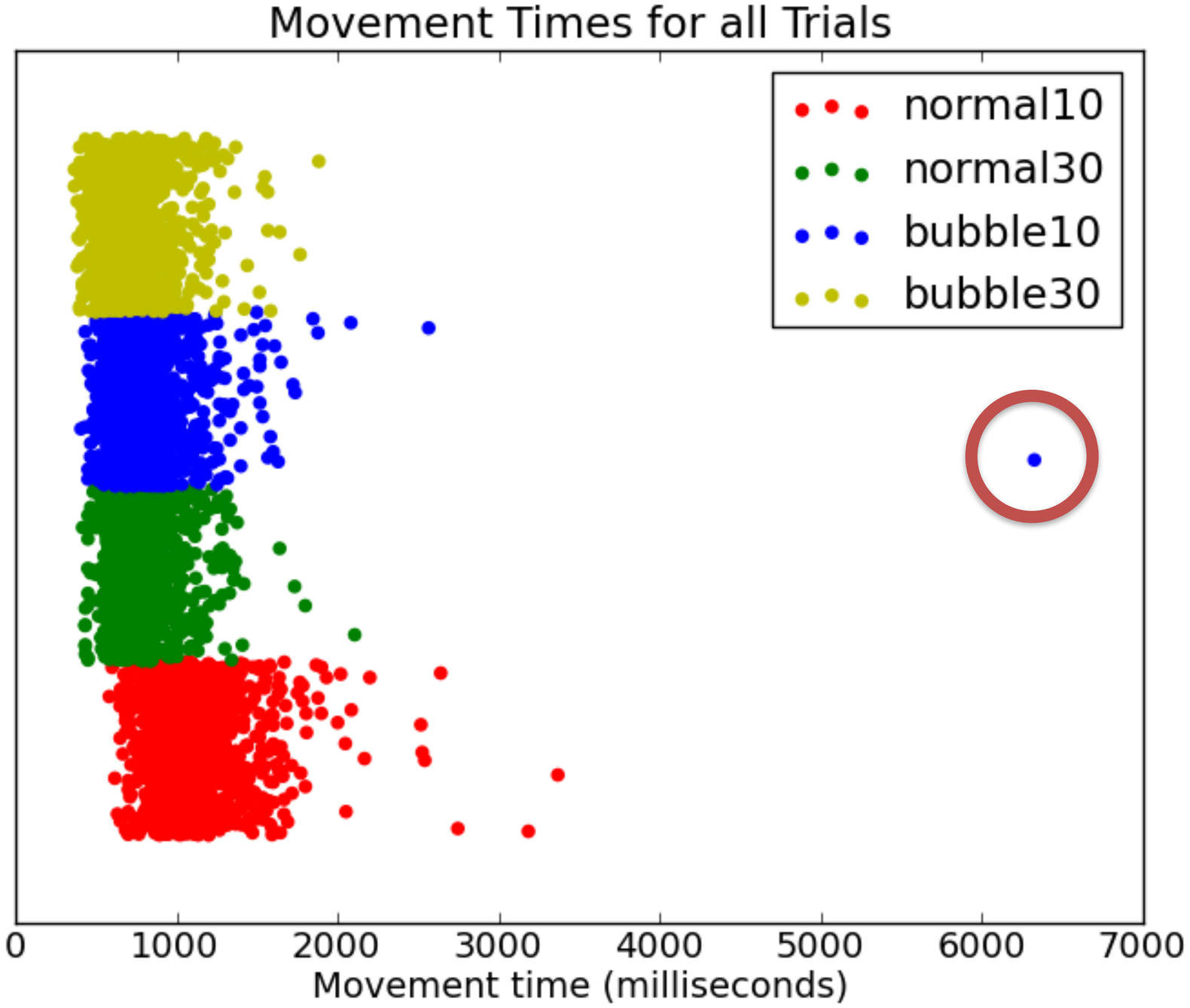
Use different visualizations: Histograms, scatterplots, box plots, ...

1D Scatter Plot



1D Scatter Plot ▪ Colored by Condition



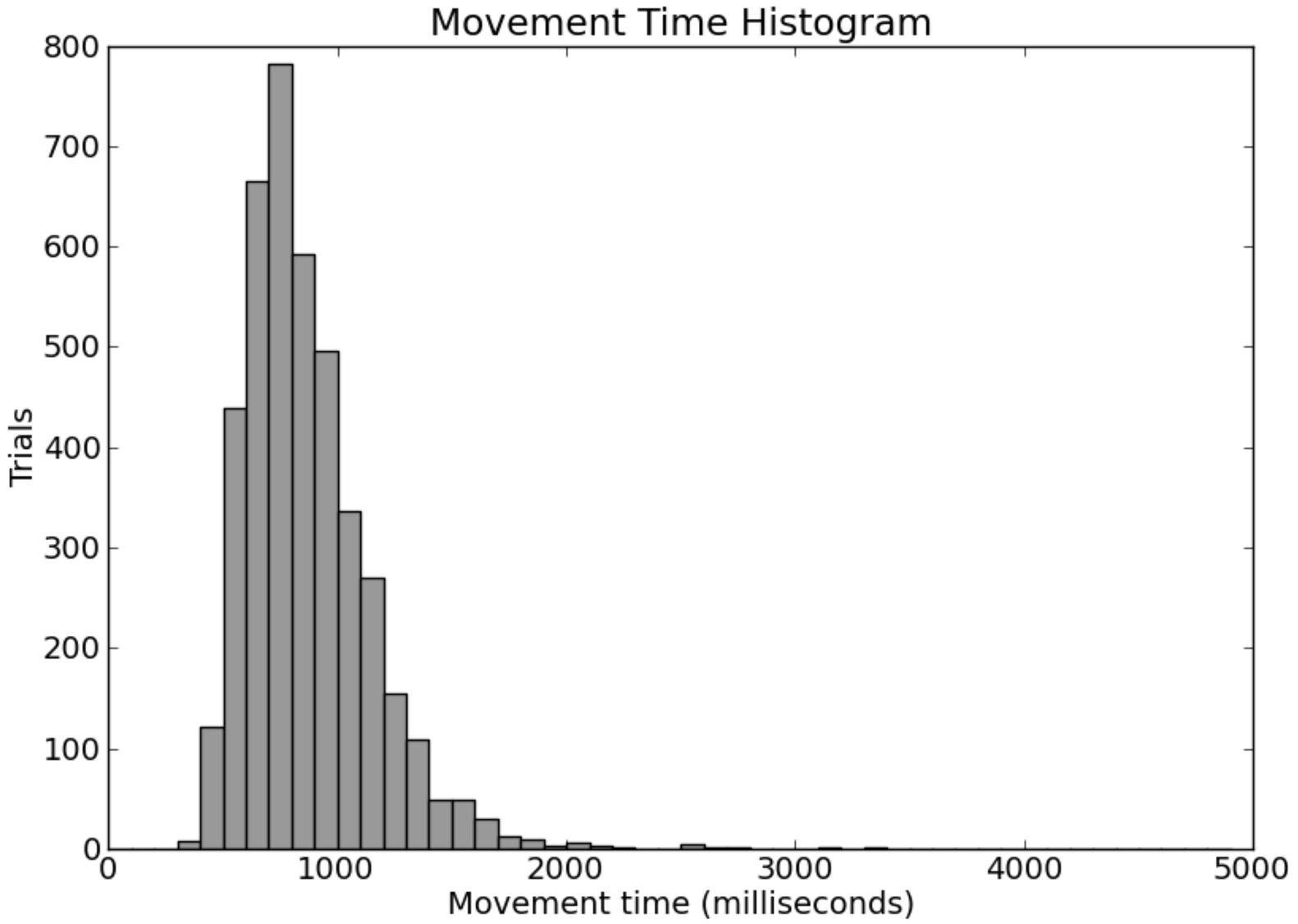

Movement Times for all Trials
normal10
normal30
bubble10
bubble30
0
1000
2000
3000
4000
5000
6000
7000
Movement time (milliseconds)

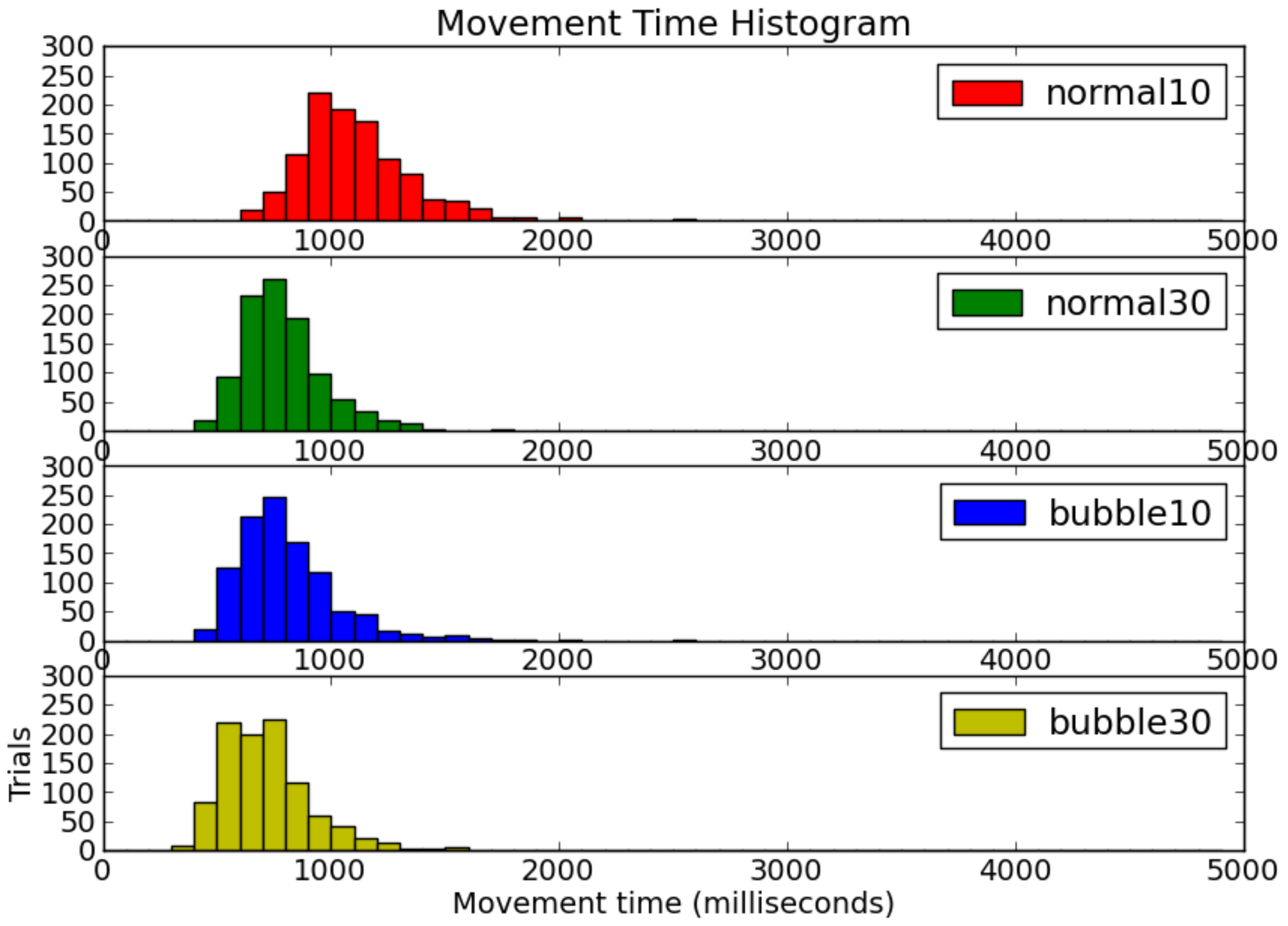
CLEANING DATA

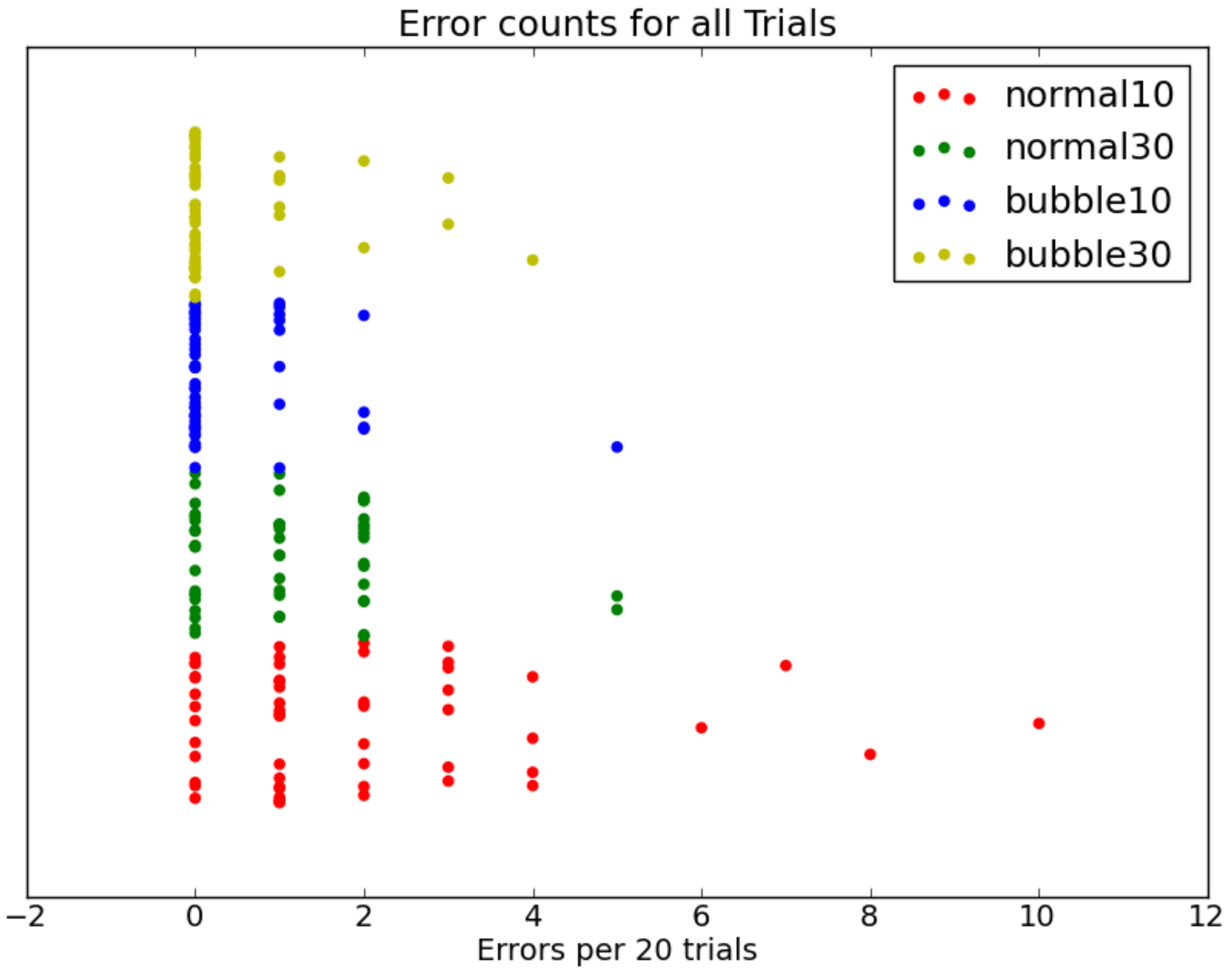
Don’t discard data just because it doesn’t fit your expectation!

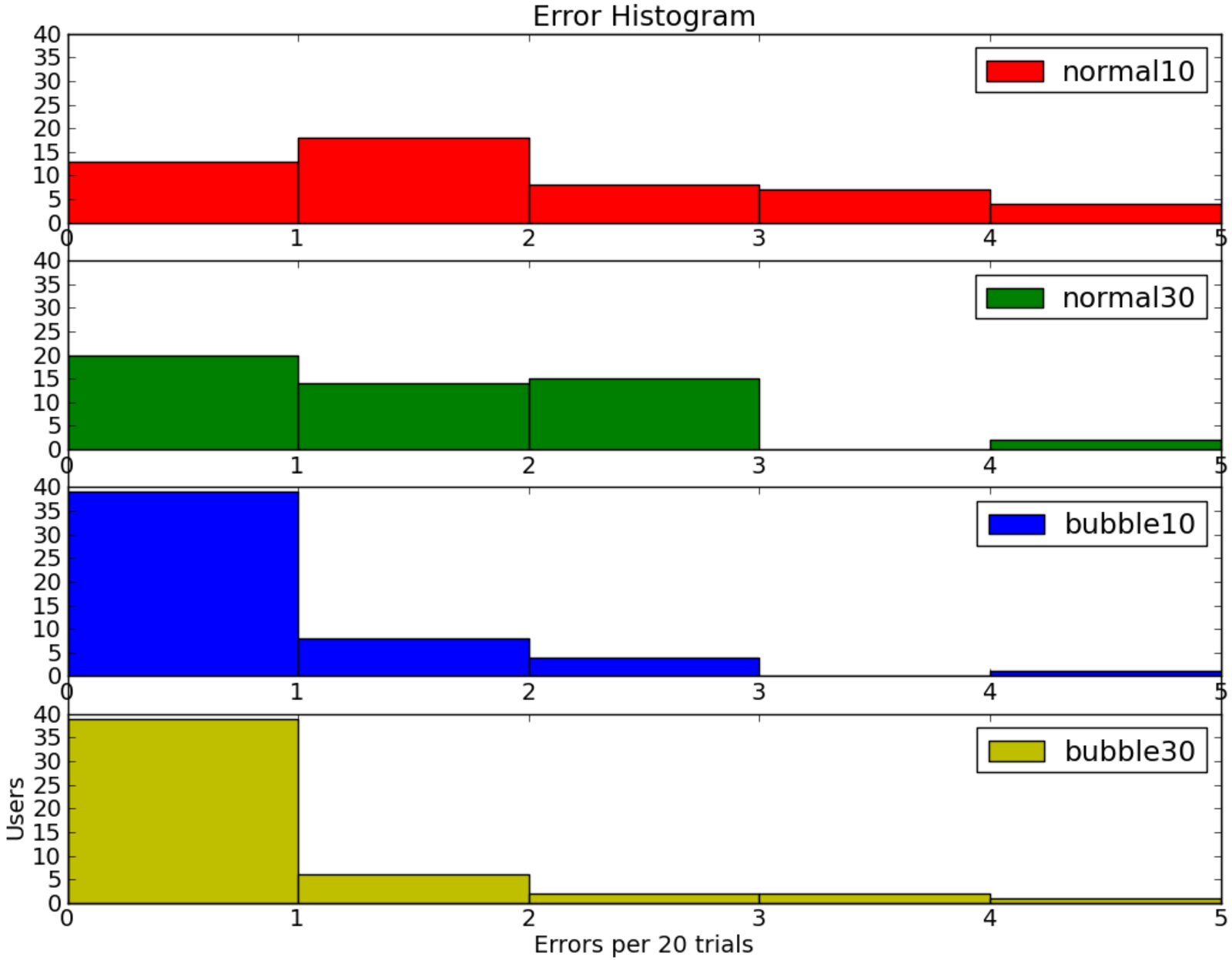
Maybe your assumptions were wrong.

In online experiments, discarding extreme outliers can make sense if you believe they reflect users not following normal task protocol (e.g., multitasking in a reaction-time study)


Movement Time Histogram
Trials
Movement time (milliseconds)
0
100
200
300
400
500
600
700
800
0
1000
2000
3000
4000
5000


Movement Time Histogram
normal10
normal30
bubble10
bubble30
Trials
Movement time (milliseconds)
0
50
100
150
200
250
300
1000
2000
3000
4000
5000


Error counts for all Trials
normal10
normal30
bubble10
bubble30
−2
0
2
4
6
8
10
12
Errors per 20 trials


Error Histogram
normal10
normal30
bubble10
bubble30
Users
Errors per 20 trials

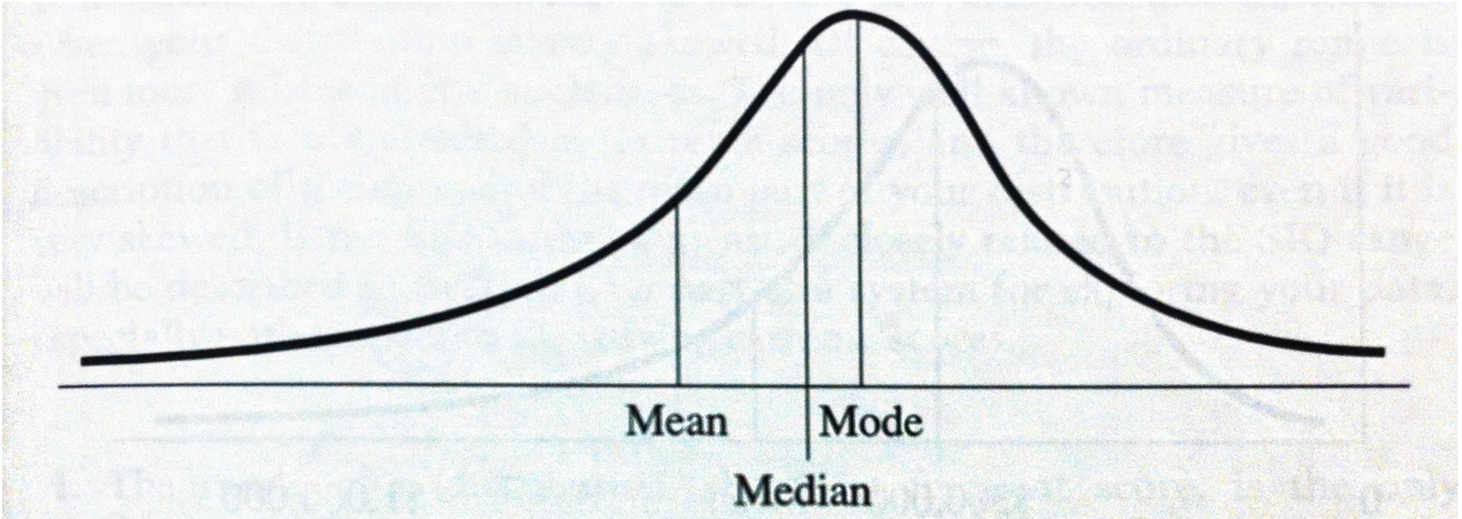
MEDIAN VS. MEAN

For **normally distributed data**, mean=median.

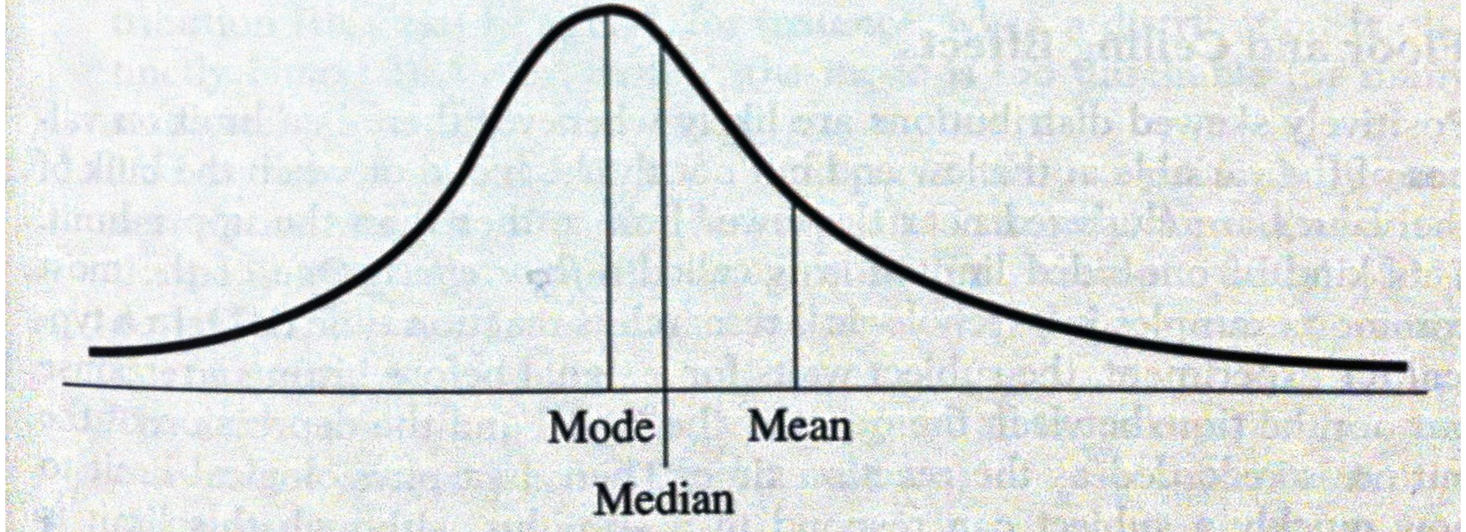
Many data sets gathered online are strongly skewed (they exhibit power law distributions – “long tails”)

Outliers pull the mean to the right/left

Median is more robust!



a. Negatively skewed distribution



b. Positively skewed distribution

POWER LAW DISTRIBUTIONS

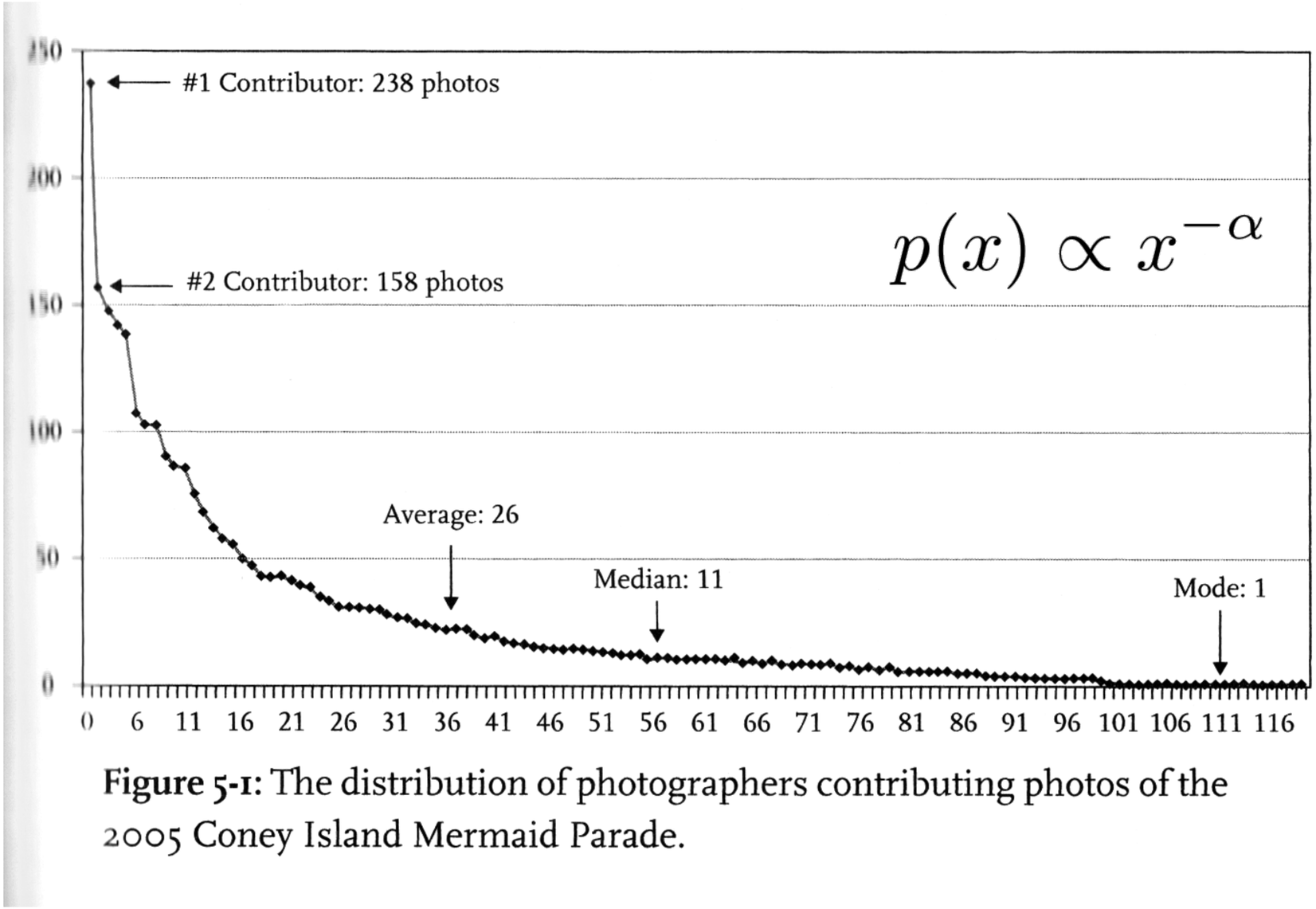
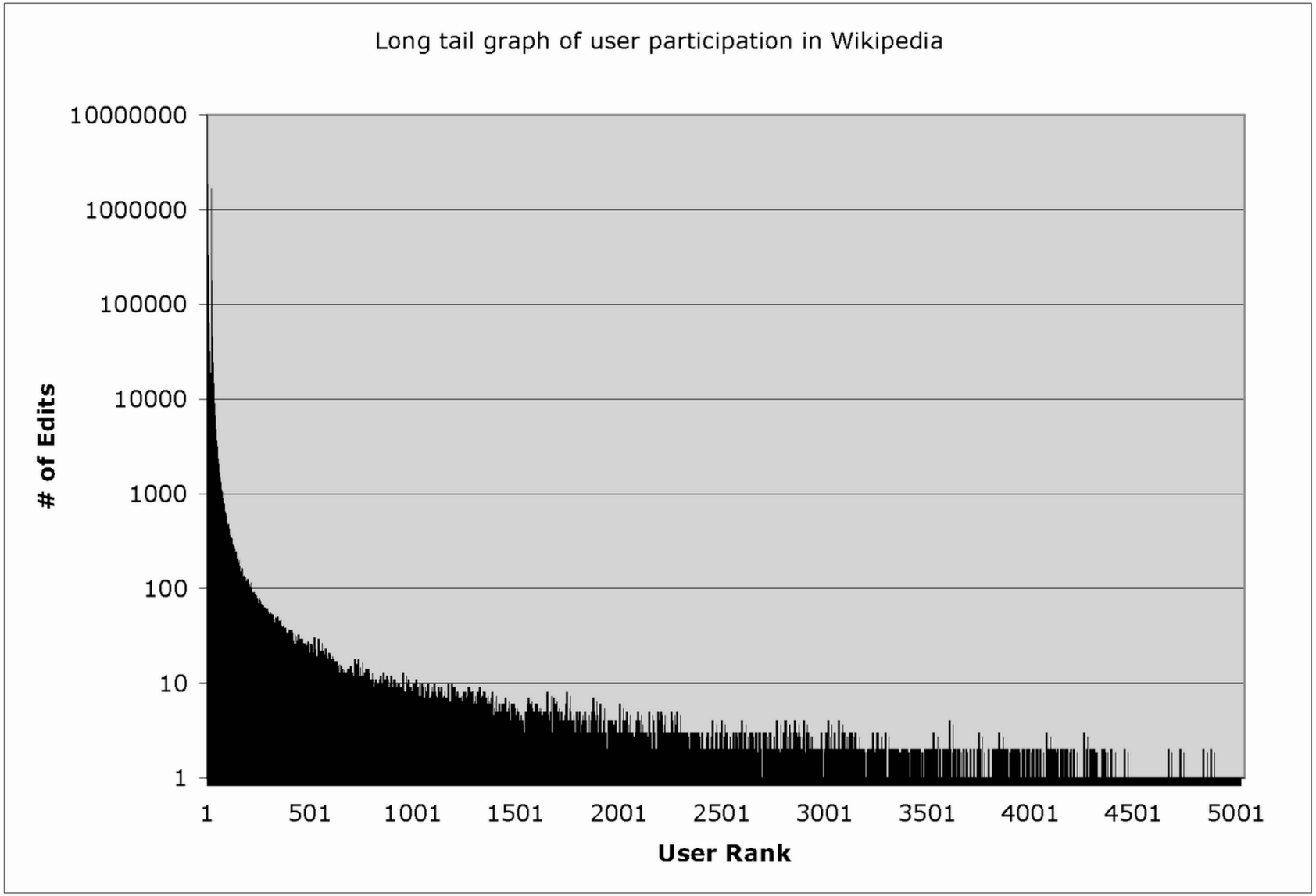
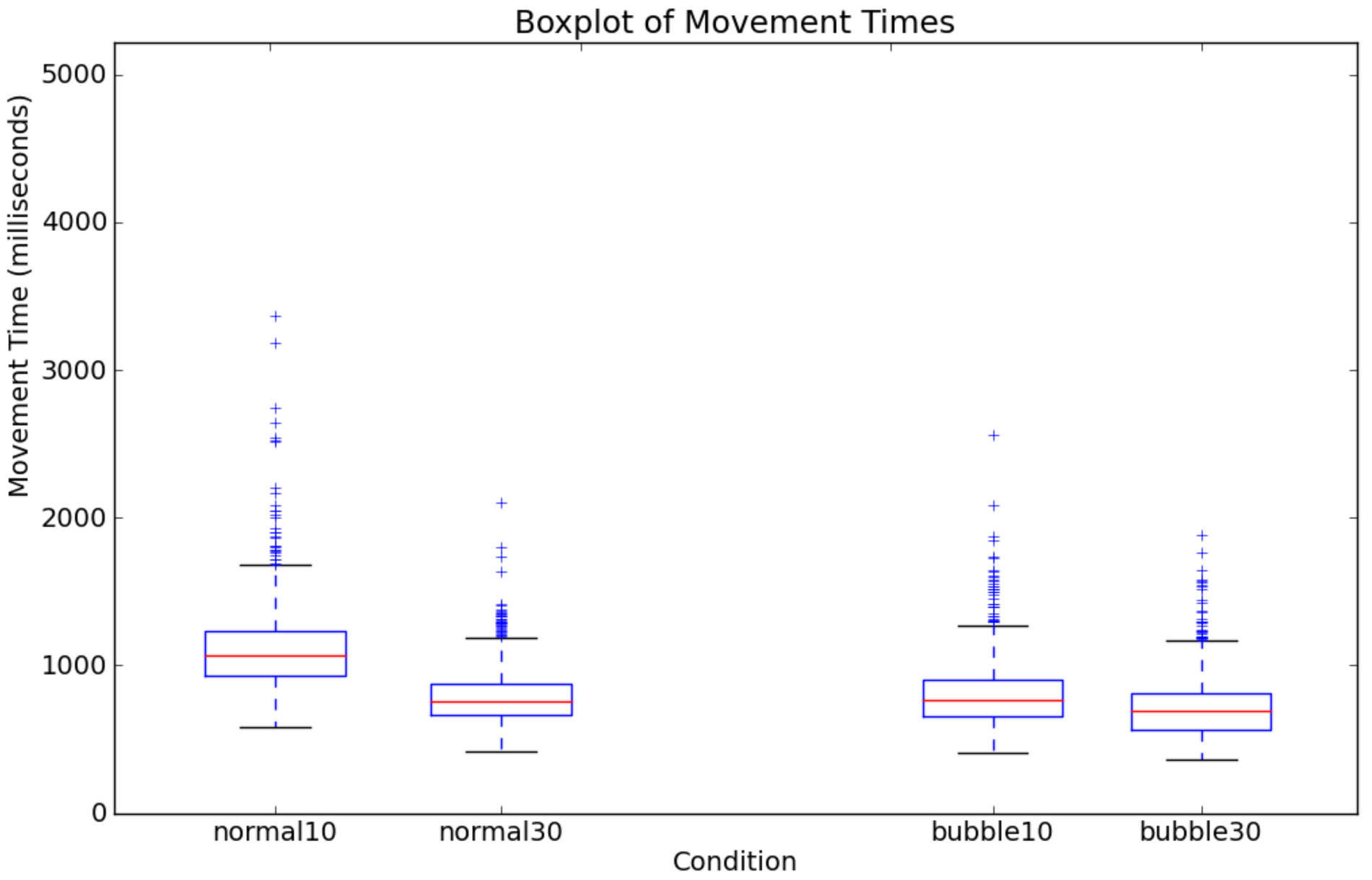
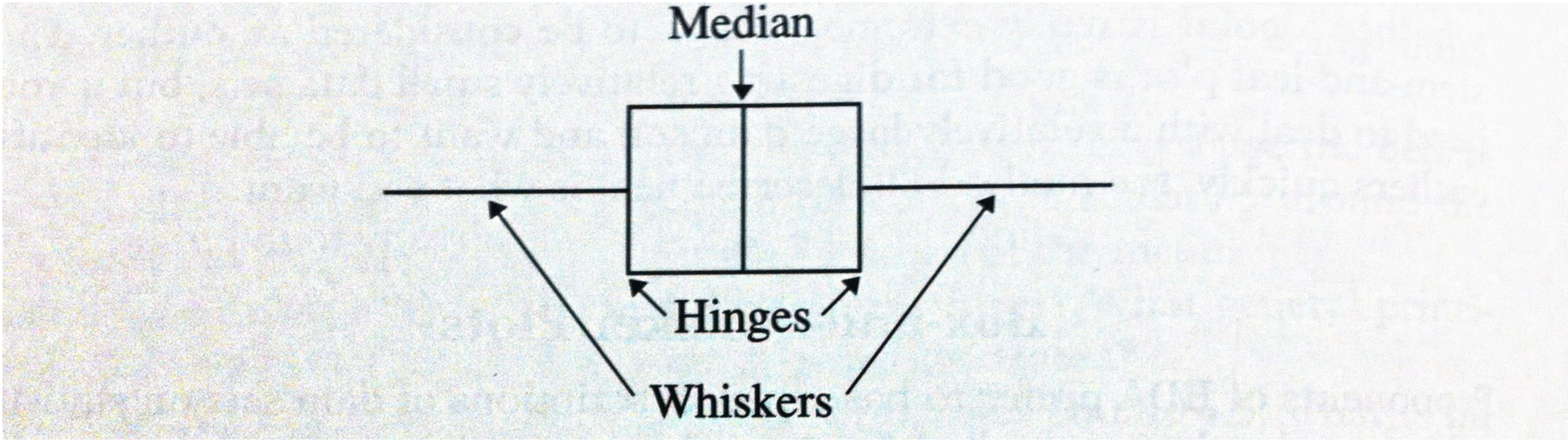


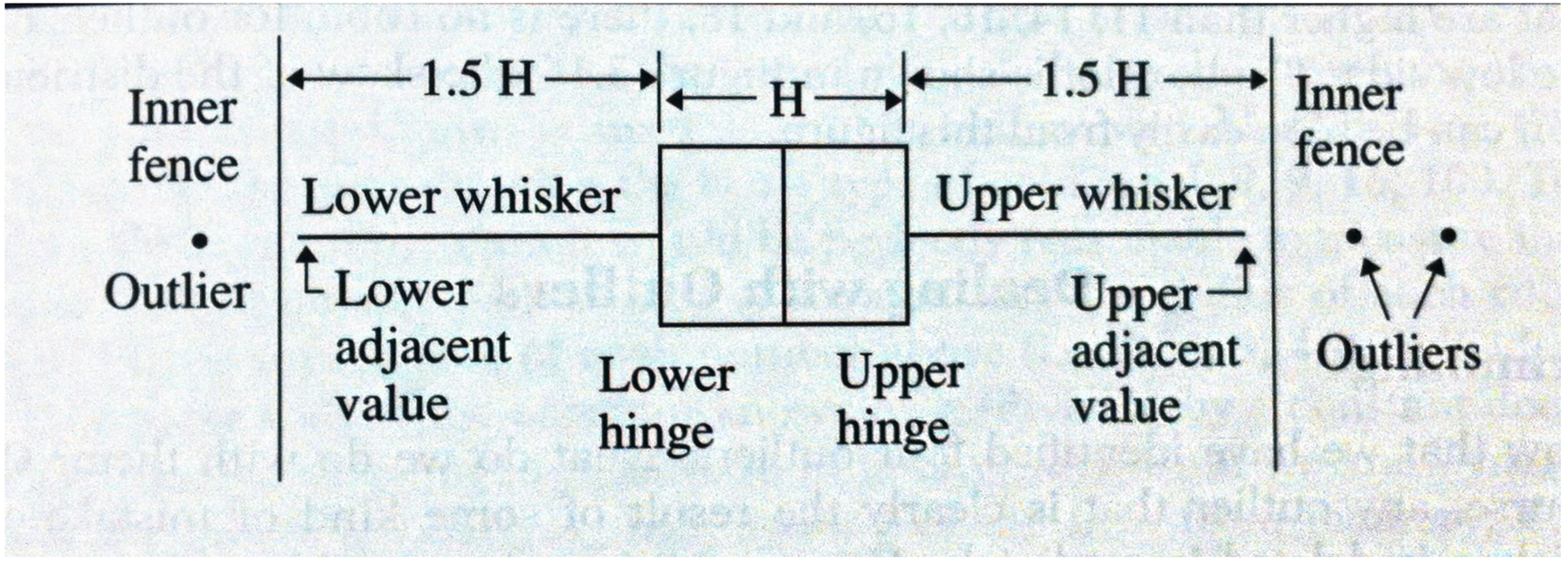
Figure 5-1: The distribution of photographers contributing photos of the 2005 Coney Island Mermaid Parade.

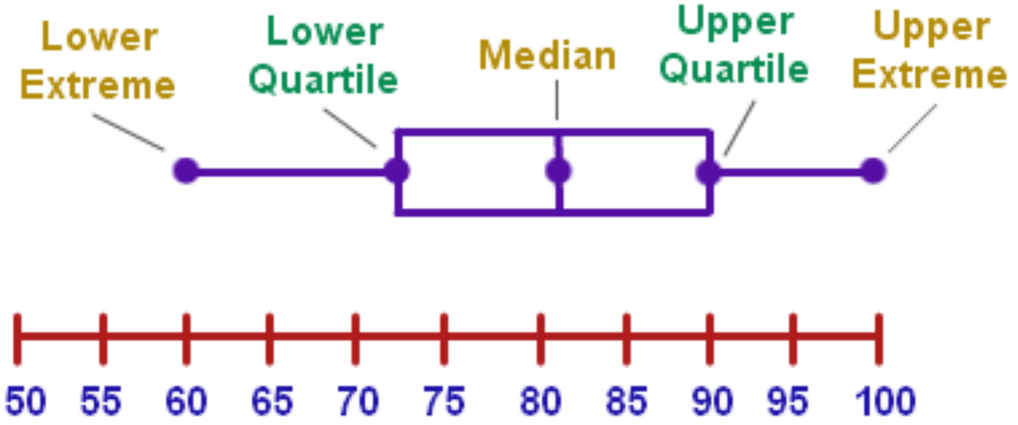
POWER LAW DISTRIBUTION

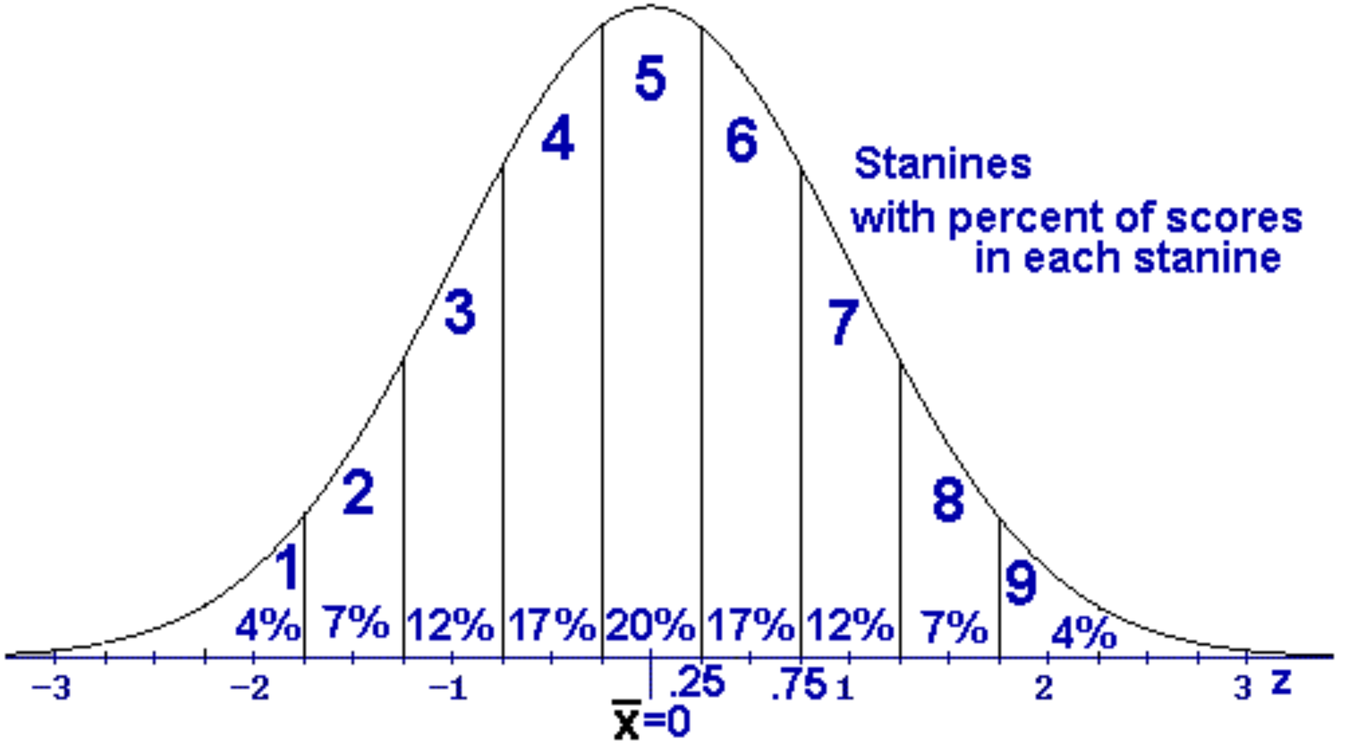



Boxplot of Movement Times
Movement Time (milliseconds)
5000
4000
3000
2000
1000
0
normal10
normal30
bubble10
bubble30
Condition

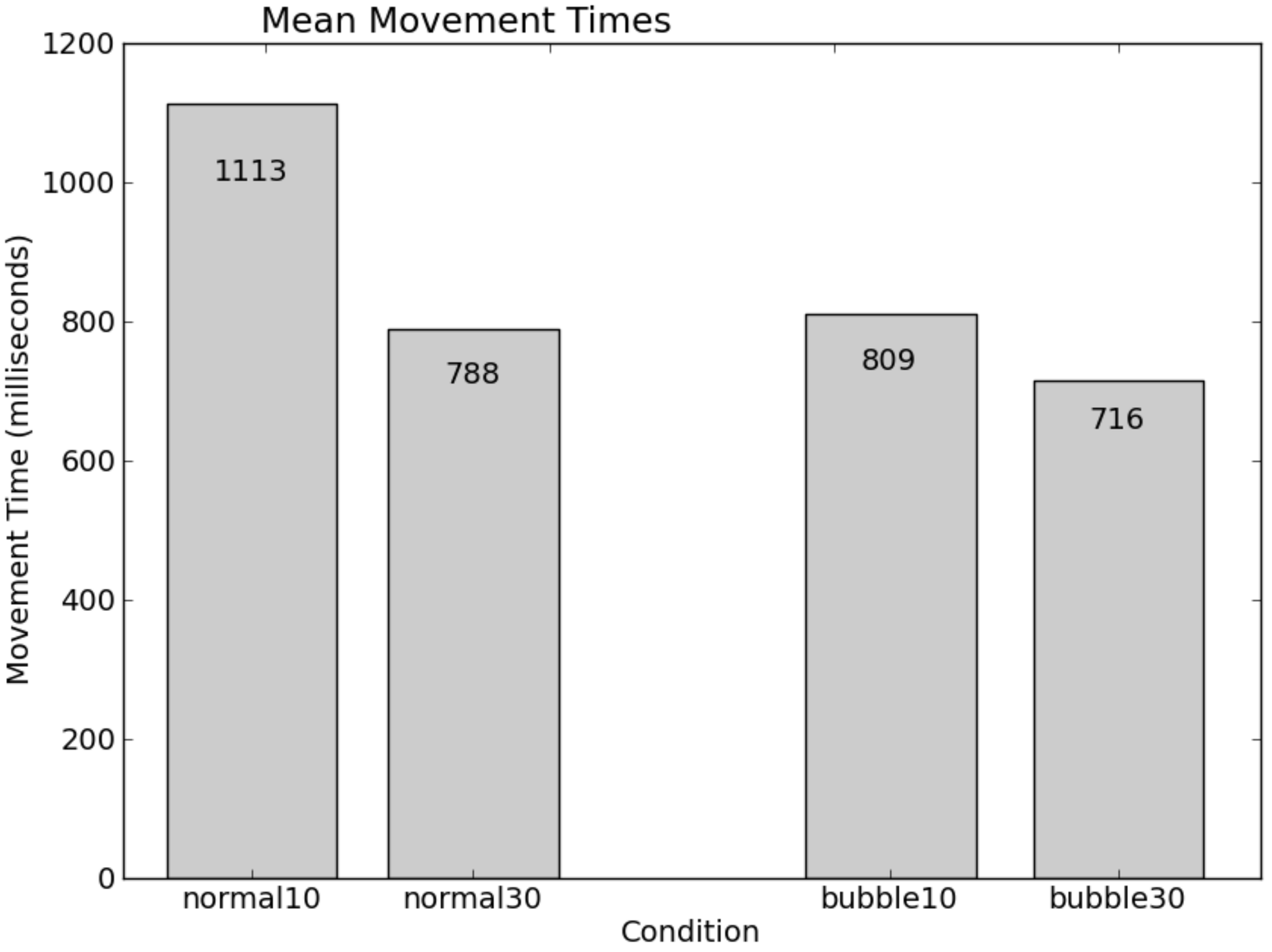

Median
Hinges
Whiskers

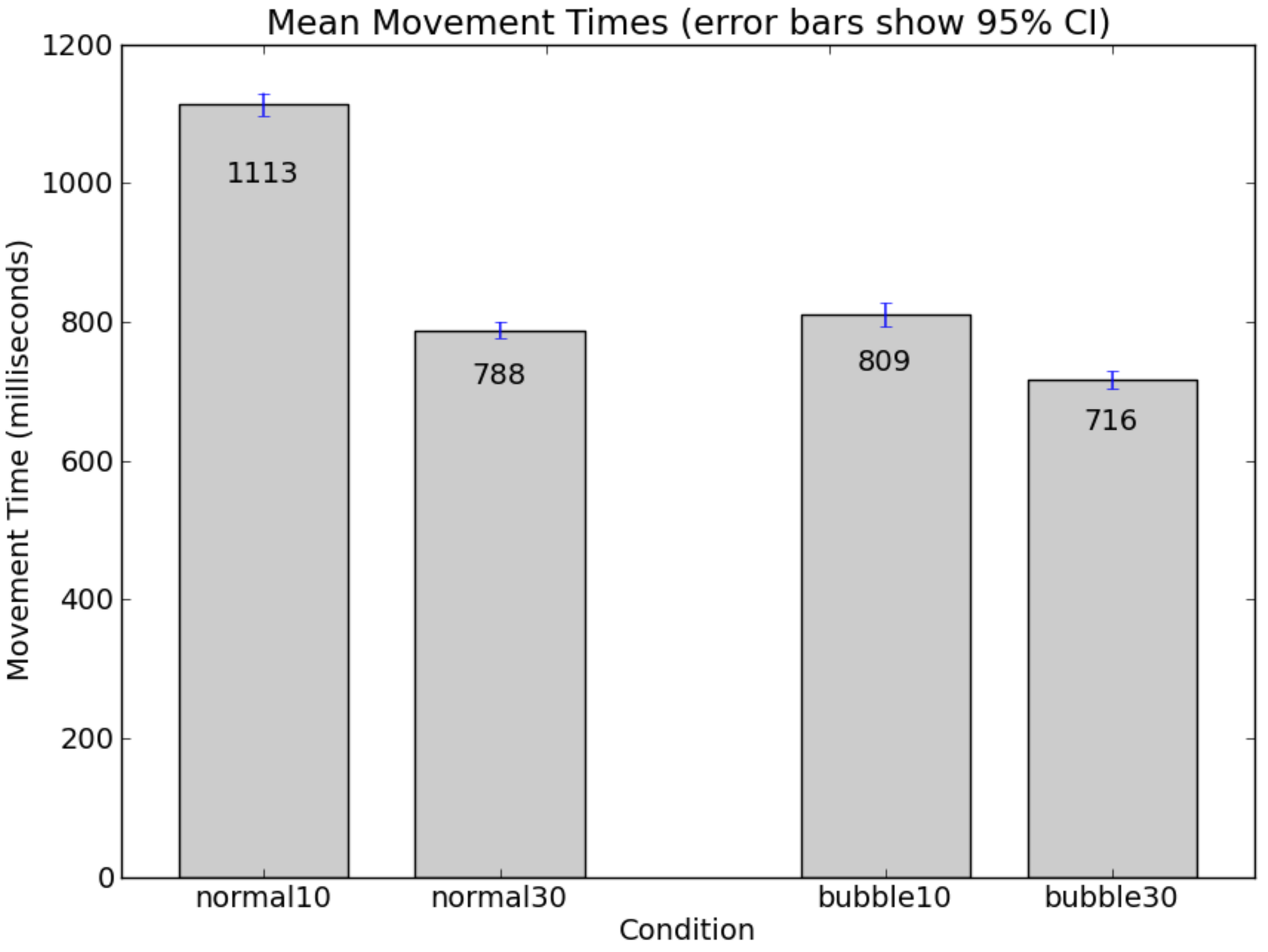

Inner fence
1.5 H
H
1.5 H
Inner fence
Lower whisker
Upper whisker
Outlier
Lower adjacent value
Upper adjacent value
Lower hinge
Upper hinge
Outliers


Lower Extreme
Lower Quartile
Median
Upper Quartile
Upper Extreme
50 55 60 65 70 75 80 85 90 95 100


Stanines
with percent of scores
in each stanine
1
2
3
4
5
6
7
8
9
4%
7%
12%
17%
20%
17%
12%
7%
4%
−3
−2
−1
.25
.75
1
2
3
z
x̄=0

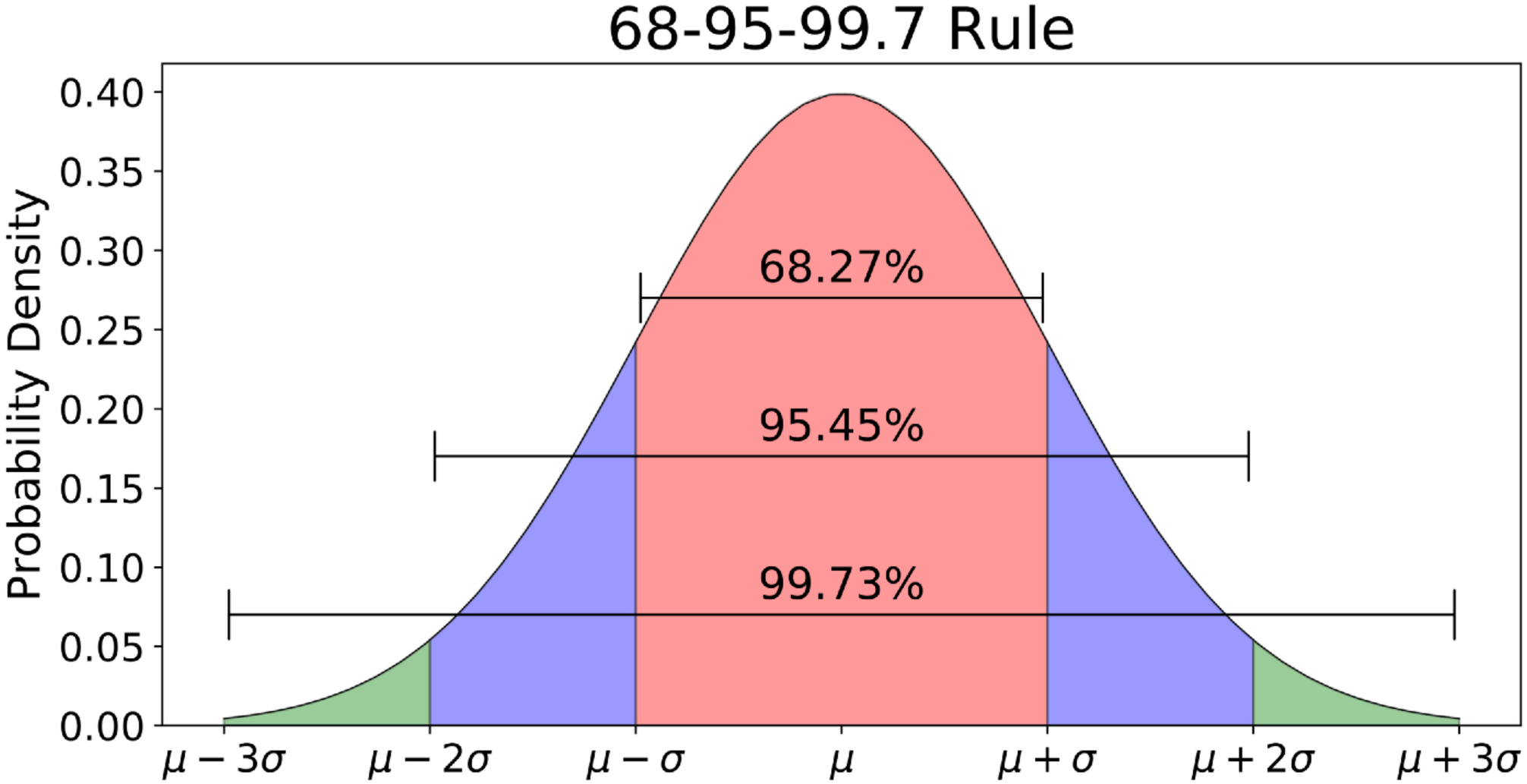
An appropriate scale
Maximum value
75th percentile
50th percentile (median)
Mean (optional)
25th percentile
Minimum value


Mean Movement Times
Movement Time (milliseconds)
1200
1000
800
600
400
200
0
1113
788
809
716
normal10
normal30
bubble10
bubble30
Condition


Mean Movement Times (error bars show 95% CI)
Movement Time (milliseconds)
1200
1000
800
600
400
200
0
1113
788
809
716
normal10
normal30
bubble10
bubble30
Condition

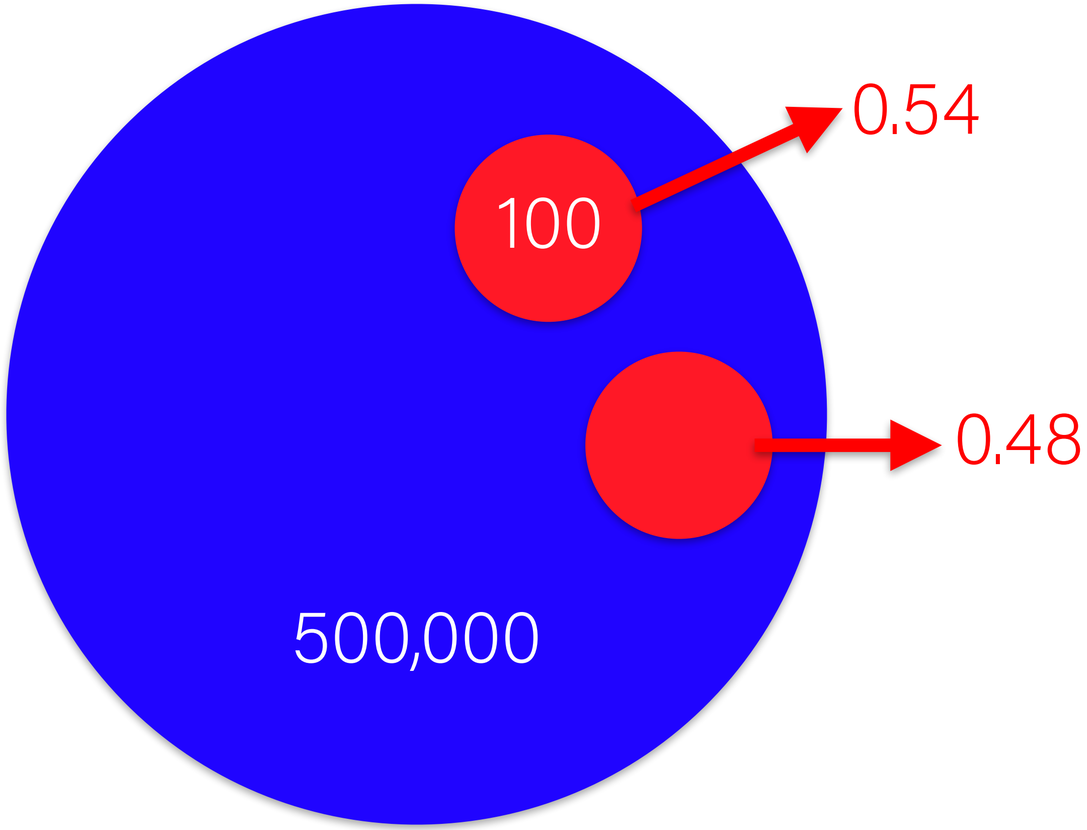
STANDARD DEVIATION

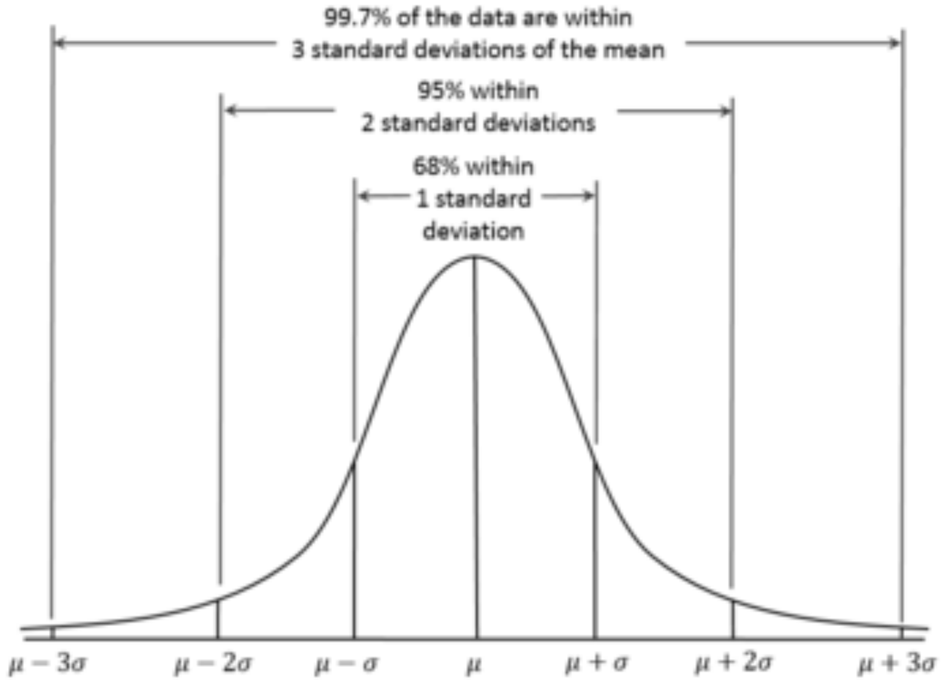
Measure of spread



CONFIDENCE INTERVAL

Candidate A or Candidate B





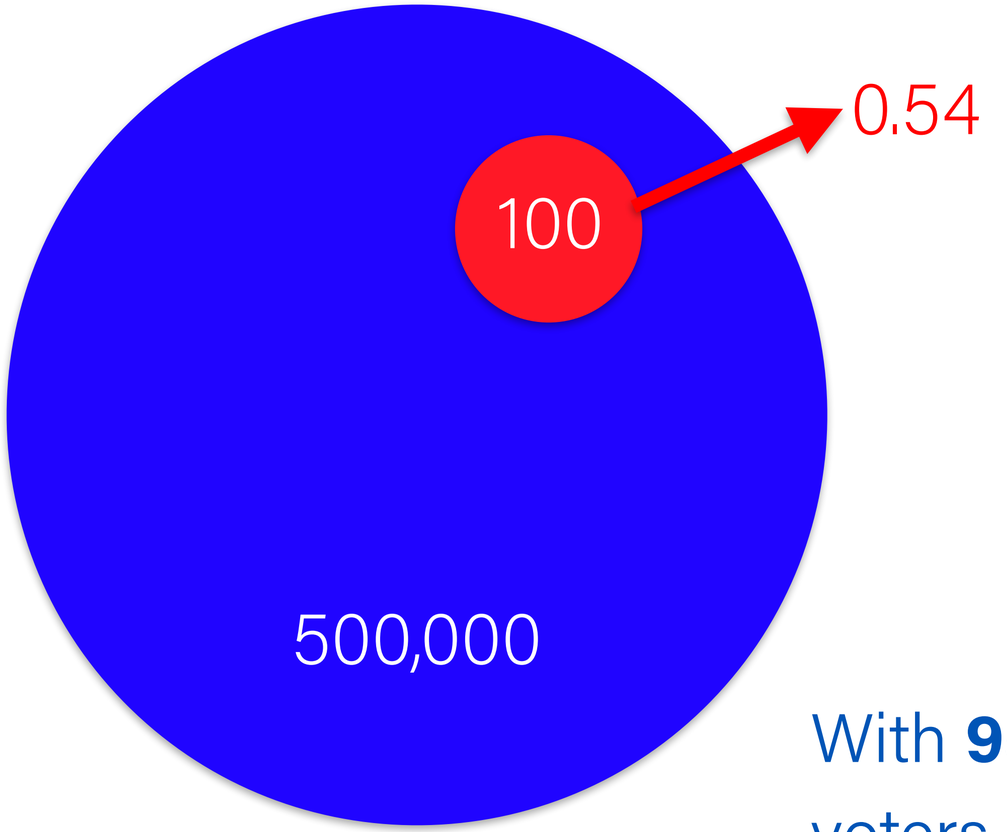
95% chance our sample proportion is within is within 2 standard deviations of true proportion

There is a 95% chance that our true proportion is within 2 standard deviations of our sample proportion

CONFIDENCE INTERVAL

Candidate A or Candidate B

$$SE = \sqrt{\frac{p(1-p)}{n}}$$



Standard error of sample proportion

p = sample proportion **0.54**

n = sample size **100**

SE ≈ **0.05**

Confidence Level

With **95% confidence** between **0.44 and 0.64** of voters support candidate A

Confidence Interval

Margin of Error — since we care about 95% confidence need two standard errors on each side 0.05 X 2 = 0.10 **±0.1**

CONFIDENCE INTERVAL

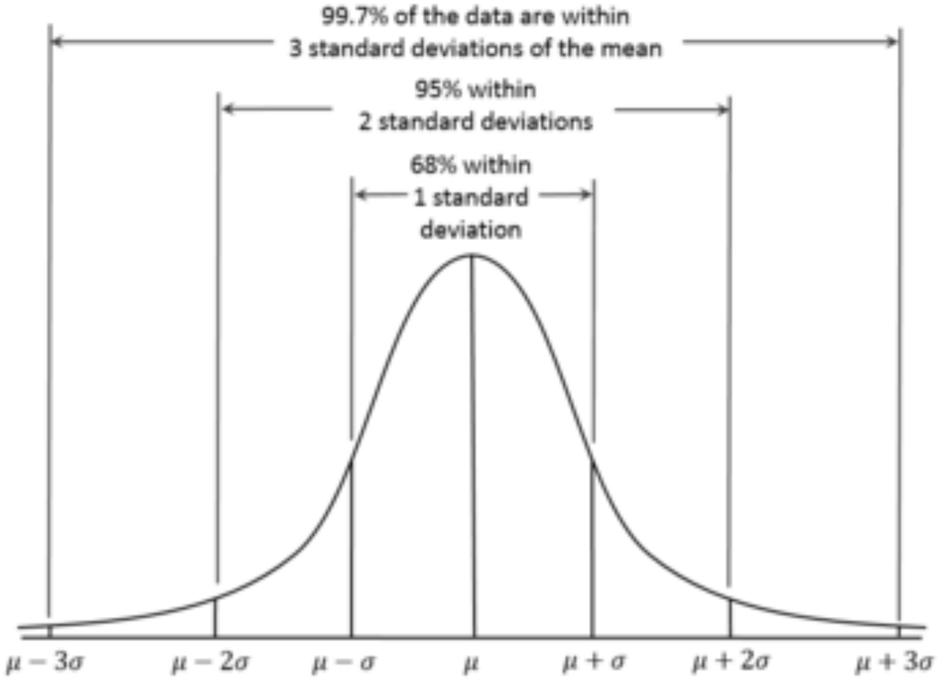
confidence interval (also called margin of error) is the plus-or-minus figure usually reported in newspaper or television opinion poll results.

For example, if you use a confidence interval of 4 and 47% percent of your sample picks an answer you can be "sure" that if you had asked the question of the entire relevant population between 43% (47-4) and 51% (47+4) would have picked that answer

CONFIDENCE LEVEL

confidence level tells you how sure you can be expressed as a percentage and represents how often the true percentage of the population who would pick an answer lies within the confidence interval.

The 95% confidence level means you can be 95% certain



CONFIDENCE LEVEL

Confidence level refers to the percentage of all possible samples that can be expected to include the true population parameter. For example, suppose all possible samples were selected from the same population, and a confidence interval were computed for each sample.

A 95% confidence level implies that 95% of the confidence intervals would include the true population parameter.

SAMPLE SIZE

1000 people in population

95% confidence level

Confidence interval of **±5**

Need to sample **278** people

Confidence interval of **±1**

...you need to sample **906** people

1000 people in population

99% confidence level

Confidence interval of **±5**

Need to sample **400** people

Confidence interval of **±1**

...you need to sample **943** people

EFFECT SIZES: TIME

Normal vs. Bubble cursor at target size 10:
1113ms vs. 810ms: **Bubble cursor 27% faster**

Normal vs. Bubble cursor at target size 30:
788ms vs. 716ms: **Bubble cursor 9% faster**

Target size for normal cursor:
1113ms vs 788ms: **Larger targets 29% faster**

Target size for Bubble cursor:
810ms vs. 716ms: **Larger targets 11% faster**

EFFECT SIZES: ERROR

Normal vs. Bubble cursor, target size 10:

1.89 vs. 0.4 Errors per 20 trials: **79% fewer errors**

Normal vs. Bubble cursor, target size 30:

1.06 vs. 0.02 Errors per 20 trials: **98% fewer errors**

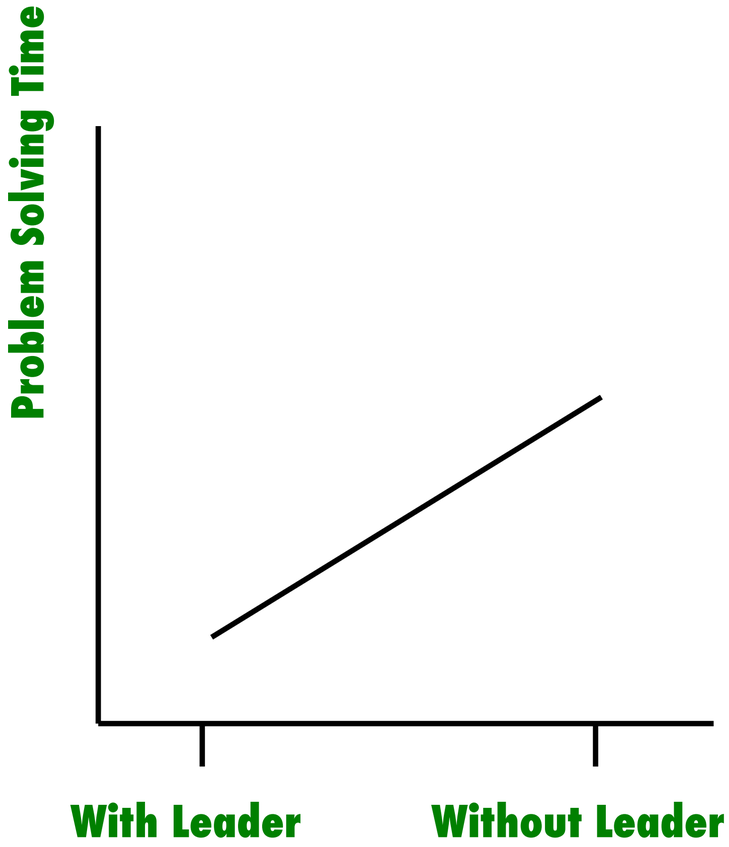
INTERACTION EFFECTS

Relationship between one IV and DV depends on the level of another IV

EXAMPLE OF INTERACTIONS

Group problem solving

Independent variable: Leadership

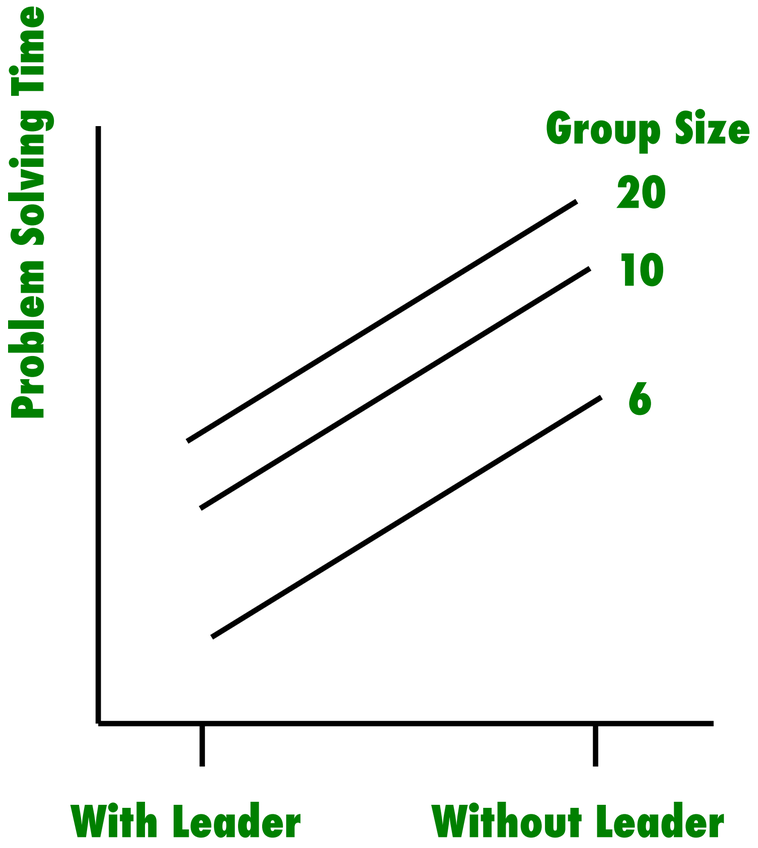


EXAMPLE OF INTERACTIONS

Group problem solving

Independent variable: Leadership

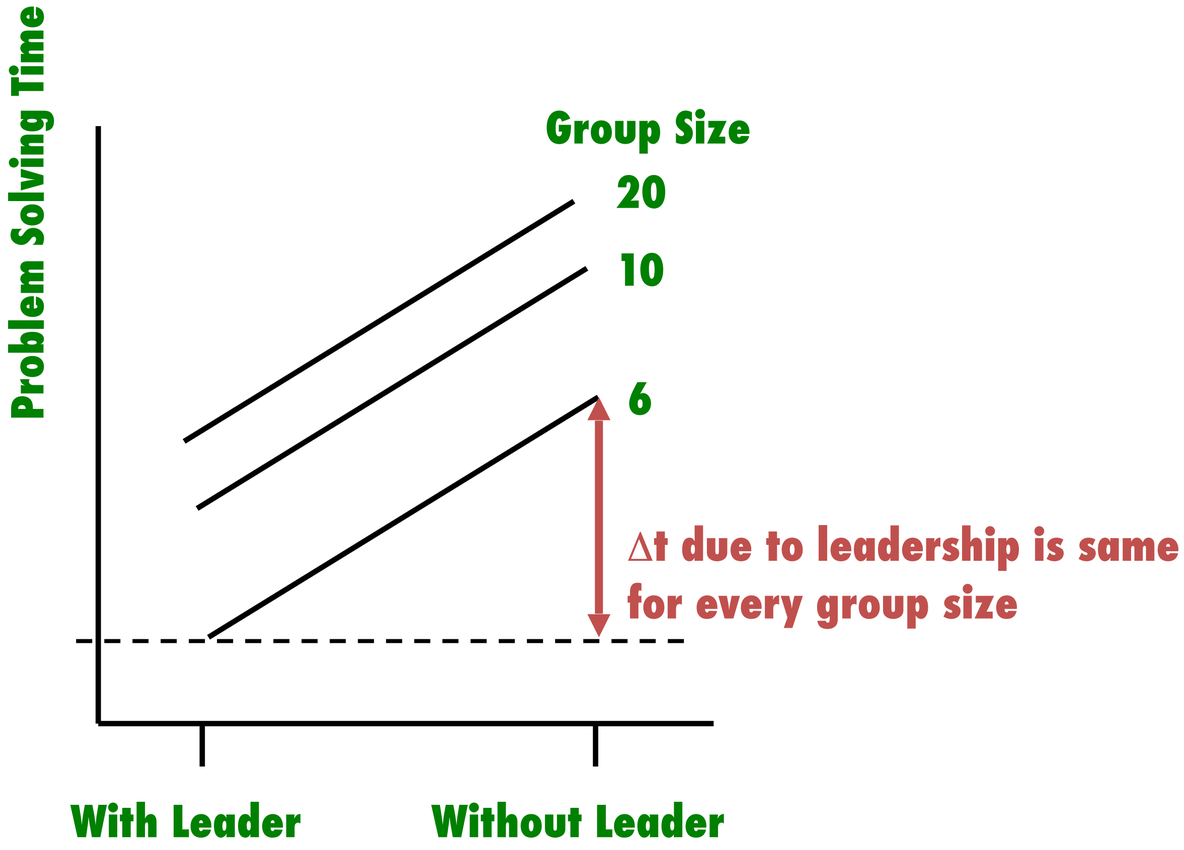
Independent variable: Group size



EXAMPLE OF INTERACTIONS

Group problem solving

Change in time due to leadership is same regardless of group size



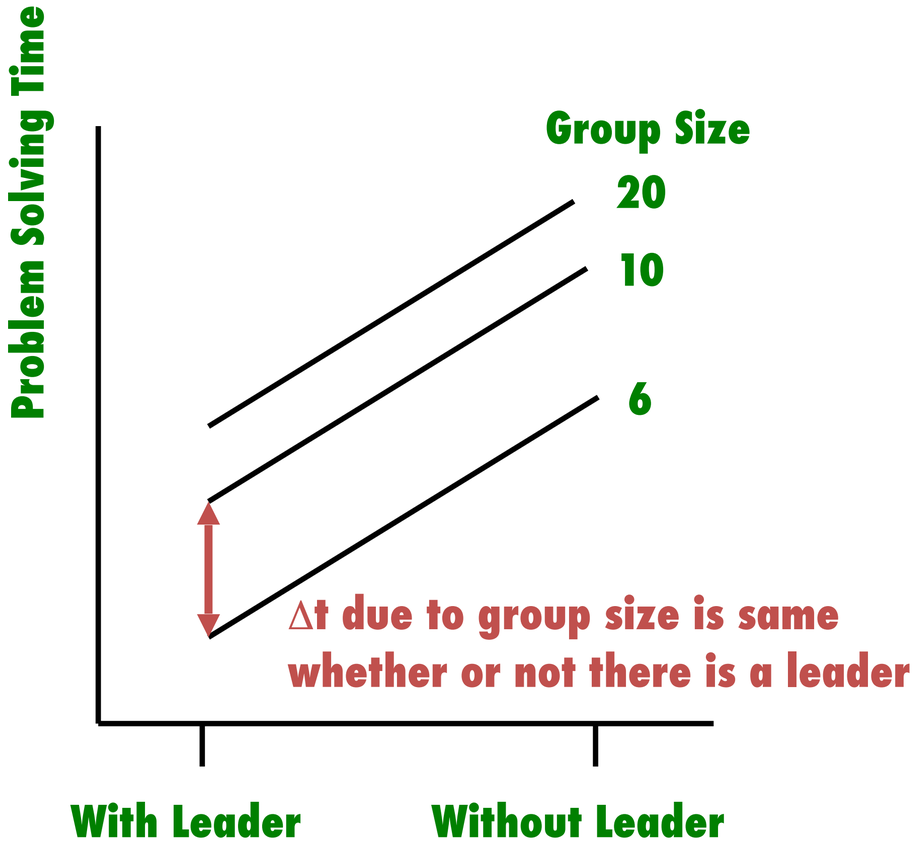
EXAMPLE OF INTERACTIONS

Group problem solving

Change in time due to leadership is same regardless of group size

Change in time due to group size is same regardless of leadership

Independent variables do not interact

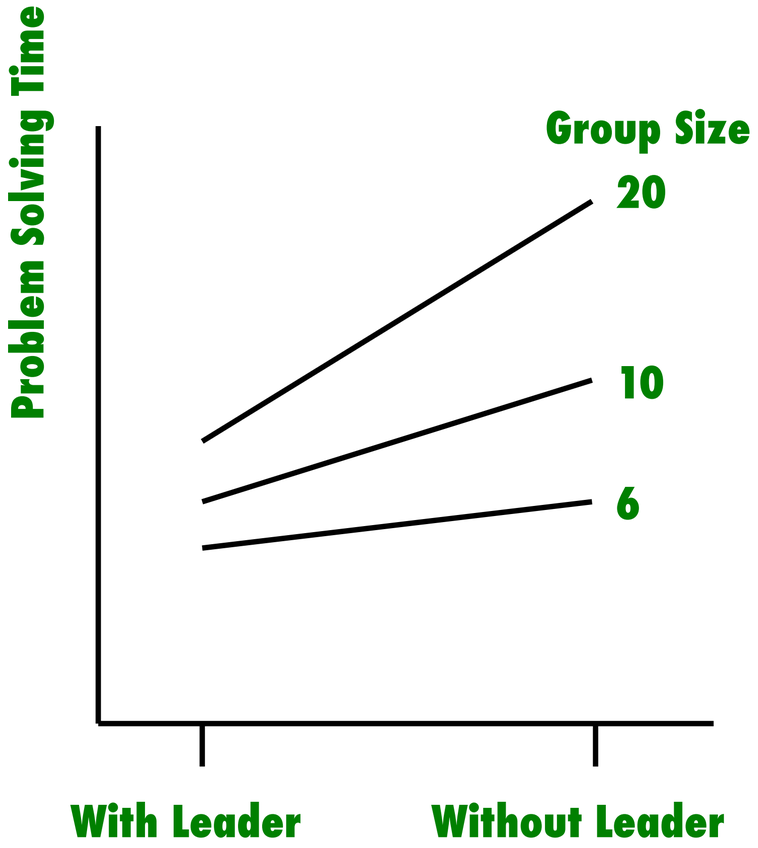


EXAMPLE OF INTERACTIONS

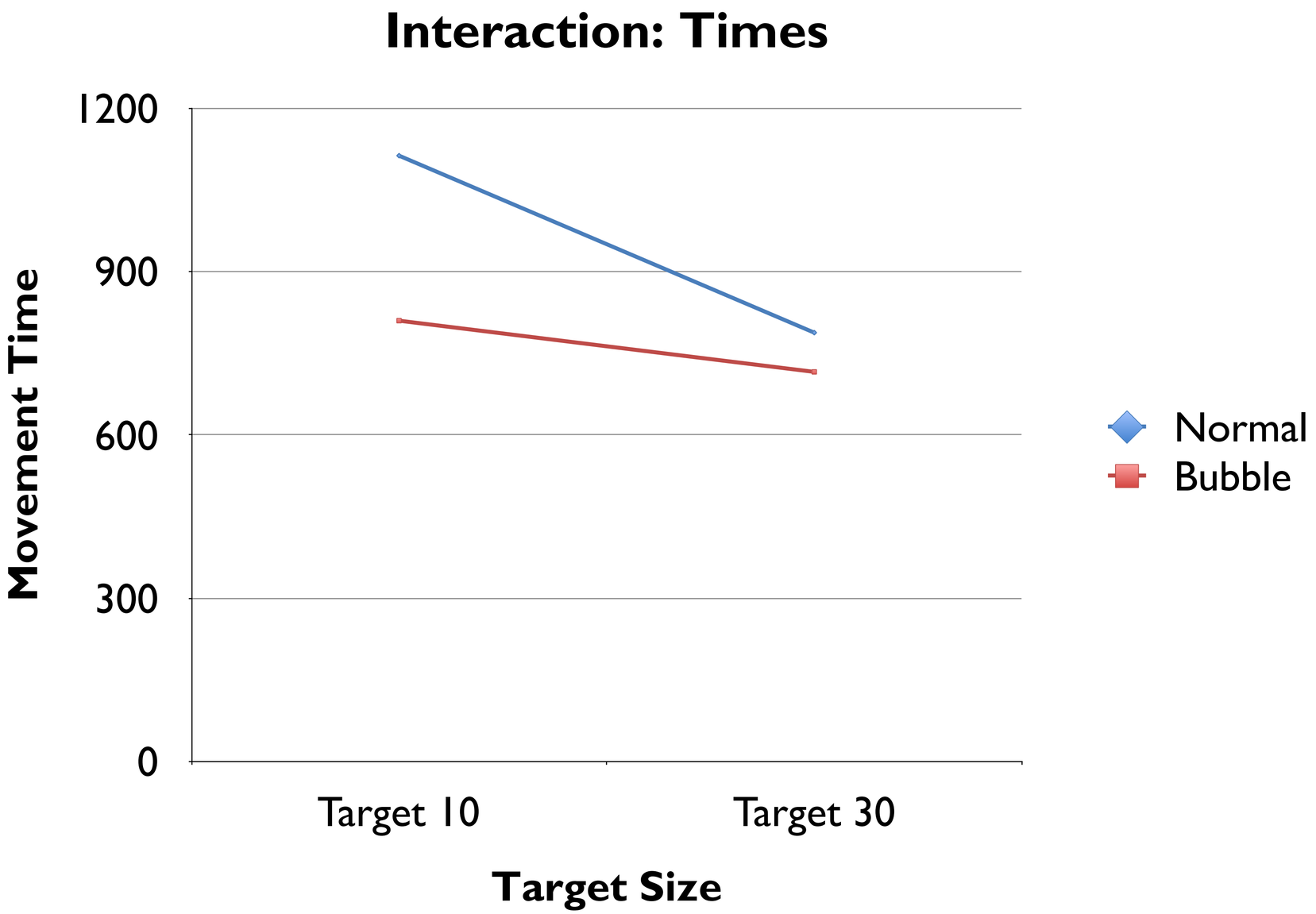
Multiple IVs effect DV non-additively

Change in time due to leadership differs with changes in group size

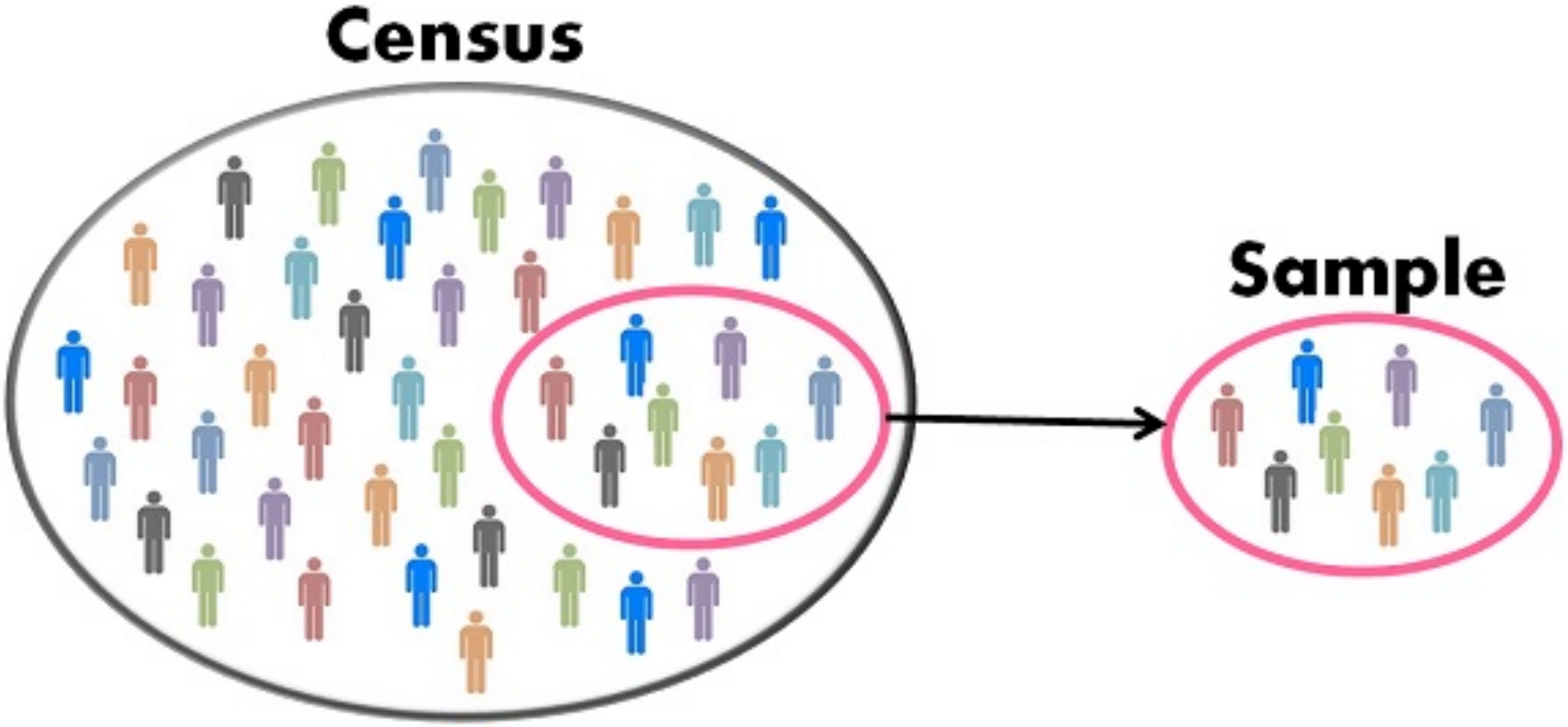
Independent variables do interact




Interaction: Errors
Errors per 20 trials
2
1.5
1
0.5
0
Target 10
Target 30
Target Size
Normal
Bubble


Interaction: Times
1200
900
600
300
0
Movement Time
Target 10
Target 30
Target Size
Normal
Bubble

POPULATION VERSUS SAMPLE



ARE THE RESULTS MEANINGFUL?

Hypothesis testing

Hypothesis: Manipulation of IV **effects** DV in some way

Null hypothesis: Manipulation of IV has **no effect** on DV

Null hypothesis assumed true unless statistics allow us to reject it

Statistical significance (p value)

Likelihood that results are due to chance variation

$p < 0.05$ usually considered significant (Sometimes $p < 0.01$)

Means that < 5% chance that null hypothesis is true

Statistical tests

T-test

Correlation

ANOVA (1 factor, > 2 levels, multiple factors)

MANOVA (> 1 dependent variable)

ANOVA - ANALYSIS OF VARIANCE

Single factor analysis of variance (ANOVA)

Statistical method used to compare the means of 2 or more groups (one factor with two levels)

Multi-Way Analysis of variance (n-Way ANOVA)

Compare more than one independent variable

Can find interactions between independent variables

Multi-variate analysis of variance (MANOVA)

Compare between more than one dependent var.

ANOVA tests whether means differ, but does not tell us which means differ – for this we must perform pairwise t-test

OUR EXAMPLE

Two-Way ANOVA (Cursor, Size) for **time**:

Main effect for **cursor**

$F(1,4136) = 641.03$, $p<0.001$ is statistically significant.

Main effect for **size**

$F(1,4136)=778.31$, $p<0.001$ is statistically significant.

Interaction **cursor x size**

$F(1,4136)=232.94.2$, $p<0.001$ is statistically significant.

OUR EXAMPLE

Two-Way ANOVA (Cursor, Size) for **errors**:

Main effect for **cursor**

$F(1,203) = 32.4$, $p<0.001$ is statistically significant.

Main effect for **size**

$F(1,203)=4.9$, **$p=0.02$** is statistically significant.

Interaction **cursor x size**

$F(1,203)=4.7$, **$p=0.03$** is statistically significant.

ERRORS IN BUBBLE CURSOR CASE ONLY

$F(1,2038) = 0.009$, **$p=0.92$ – NOT significant**

WHAT DOES P>0.05 MEAN?

No statistically significant difference (at 5% level)

Are the two conditions thus equivalent?

NO! We DID observe differences

But can’t be sure they are not due to chance.

If the p-value is less than 0.05, we reject the null hypothesis that there's no difference between and conclude that a significant difference does exist

If the p-value is larger than 0.05, we cannot conclude that a significant difference exists. It may be due to chance.

KEEP CALM AND STUDY HARD