

Survey and Analysis

INFR11158/11230 Usable Security and Privacy

Dr. Jingjie Li

06/02/2024



THE UNIVERSITY
of EDINBURGH

Overview

- Blog
- Recap: lab study
- Think aloud
- Take-home

Good news blog structure :)

A Brief Review of the 2023 Locbit 3.0 Cyber-Attack on Bank Syariah Indonesia: Lessons for Improved Security, Trust, and Compliance

Word Count: 498

A Quick Introduction to Bank Syariah Indonesia

Bank Syariah Indonesia, commonly known as BSI, is an Islamic bank that was merged from various state-owned Islamic banks in Indonesia in February 2021, which became their founding year. BSI gained an impressive net profit of over 4 trillion in Indonesian Rupiah throughout 2022. BSI is the largest Islamic bank in Indonesia.

The Chronology

On May 8th, BSI consumers complained that they failed to access and do transactions on their accounts through the system and its services. They not only complained through consumer service but also through their social media, which became trending in Indonesia. In the evening of that day, BSI apologised and informed through social media that they were in the middle of system maintenance and experienced an error.

Bank Syariah Indonesia's Response

On May 11th, BSI finally responded publicly that they have resolved the issue and that their consumers' data is a top priority. Along with it, BSI hinted that BSI might have gotten cyber-attacked. Instead of quickly admitting to the cyber-attack, BSI gatekept the information to the point where the hackers themselves, Lockbit 3.0, stated online that they had stolen over 1.5 terabytes of data from the BSI system.

After BSI's "good" recovery response towards the cyber-attack, BSI fired their IT and Risk Director. Furthermore, BSI plans to launch a new mobile banking app with a high quality that will beat its competitors. Their actions reflect the statement made by Sasse et al in their research paper:

"Safety and security protect the interests of both individuals and organisations in the long term. However, they also compete for resources with fundamental tasks in the short term, and are likely to be sacrificed for short-term gain" [1].

Furthermore, BSI has failed to adhere to the Indonesian Law Number 8 the Year 1999 regarding Consumer Protection, where consumers have the right to know the truth of all definite information and they have the right to sue for damages. In this situation, the compromised data and BSI's delayed notification to its consumers, along with the denial of their right to timely and transparent information, suggest that BSI might have concealed the problem indefinitely if Lockbit 3.0 had not disclosed it in its statement. If BSI were straightforwardly and quickly informed and assured their consumers on the first day of the cyber-attack, consumers would not have complained on social media

and made it a trend, which would lessen the stress and distrust of the consumers towards BSI.

Conclusion

In conclusion, the 2023 cyber-attack on BSI highlights the pivotal significance of BSI's dedication to transparent communication, investment in advanced security measures, and strict adherence to legal compliance. This commitment is essential for rebuilding trust and positioning BSI as a role model in establishing industry standards within the financial sector. While BSI may not have disclosed all details about the cyber-attack, it is imperative for BSI to embrace these measures to protect consumers' data, uphold their rights, and strengthen the foundation of digital finance.

References

- [1] M.A. Sasse, S. Brostoff, and D. Weirich. 2021. Transforming the 'Weakest Link' — a Human/Computer Interaction Approach to Usable and Effective Security. *BT Technology Journal* 19, 122–131. <https://doi.org/10.1023/A:1011902718709>
- [2] Kompas. 2023. *Perjalanan Kasus BSI dari Gangguan Layanan sampai Hacker Minta Tebusan*. Retrieved January 24, 2024 from <https://money.kompas.com/read/2023/05/17/072027926/perjalanan-kasus-bsi-dari-gangguan-layanan-sampai-hacker-minta-tebusan?page=all#>
- [3] Detik. 2023. Wamen BUMN Ungkap BSI kena Hack Negara Komputer Tua. Retrieved January 24, 2023 from <https://finance.detik.com/moneter/d-6756638/wamen-bumn-ungkap-bsi-kena-hack-negara-komputer-tua>
- [4] Bank Syariah. 2024. Corporate History. Retrieved January 24, 2023 from https://ir.bankbsi.co.id/corporate_history.html
- [5] Government of Republic Indonesia. 1999. Indonesian Law No 8 Year 1999 regarding Consumer Protection.
- [6] CNBC Indonesia. 2023. *Impresif BSI Cetak Laba Bersih Rp. 4.26 Triliun Sepanjang tahun 2022*. Retrieved January 24, 2024 from <https://www.cnbcindonesia.com/syariah/20230202002324-29-410222/impresif-bsi-cetak-laba-bersih-rp426-t-sepanjang-tahun-2022>

Nice paper discussion :)

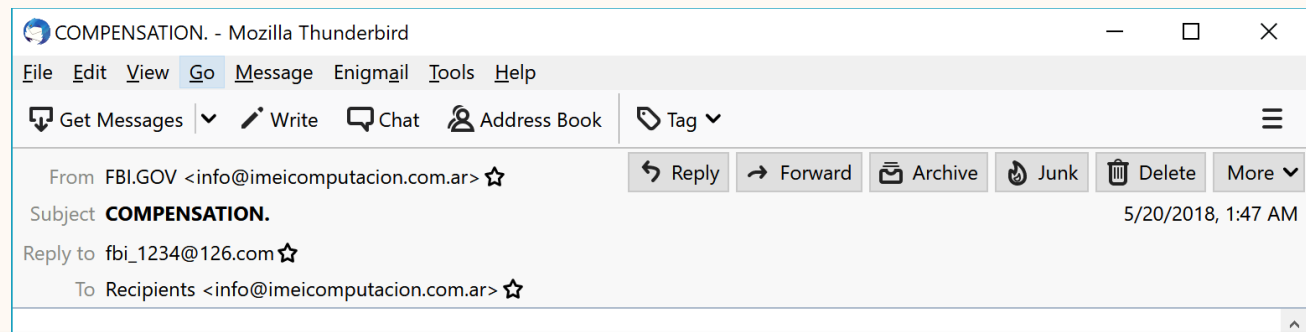
While the Nairobi-focused study provided insights, future research could be extended to other urban or rural areas in Kenya to provide a more complete picture of cybercafes use across Kenya's diverse populations and environments. Second, comparative studies with cybercafes in other developing and even developed countries could provide valuable insights into the prevalence and uniqueness of cybercafes use and security practices. Future research should ideally focus on bridging the gap between digital literacy and cybersecurity, investigating security tools developed specifically to address the needs and limitations of cybercafe users in developing countries.

A specific solution could be to design an educational course that includes basic cybersecurity concepts and common and easily overlooked issues, especially for password management, account security and identifying cyberthreats. This course is then implemented in selected cybercafes through workshops and interactive learning. The impact was subsequently assessed by means of a before-and-after comparison, which included the participants' performance on the cybersecurity knowledge test as well as their behavioural changes in reality. Feedback was then collected on the effectiveness of the course and user acceptance for improvement.

In the case of Kenya, I think perhaps a hardware-based solution is worth considering. For different scenarios such as communication accounts/gaming/social media accounts etc., mobile device-based Two-Factor Authentication should be deployed more widely to make access more difficult for malicious actors in cases where passwords can be easily obtained by them. When it comes to sensitive information such as banking that requires identity authorization, perhaps inexpensive Token-based Authentication Devices or even USB Security Keys could be considered.

Raising the security awareness of the population through education is also something that needs to be done at the same time, but educational means often rely on many other factors that do not allow the problem to be alleviated in a short period of time.

Quickly defining “phishing” so



ATM Card: We will be issuing you a custom pin based ATM card which you will use to withdraw up to \$3,000 per day from any ATM machine that has the Master Card Logo on it and the card have to be renewed in 3 years' time, which is 2021. Also with the ATM card you will be able to transfer your funds to your local bank account. The ATM card comes with a handbook or manual to enlighten you about how to use it, even if you do not have a bank account.

attempt to swindle your fund which has led to so many losses from your end and

Take note that anyone asking you for some kind of money above the usual fee is definitely a fraudster and you will have to stop communication with every other person, if you have been in contact with any. Also remember that all you will ever have to pay is \$520 U.S Dollar, and we guarantee the receipt of your fund to be successfully delivered to you in four days, after the receipt of payment has been confirmed.

with the ATM card you will be able to transfer your funds to your local bank

Questionnaires

Questionnaires

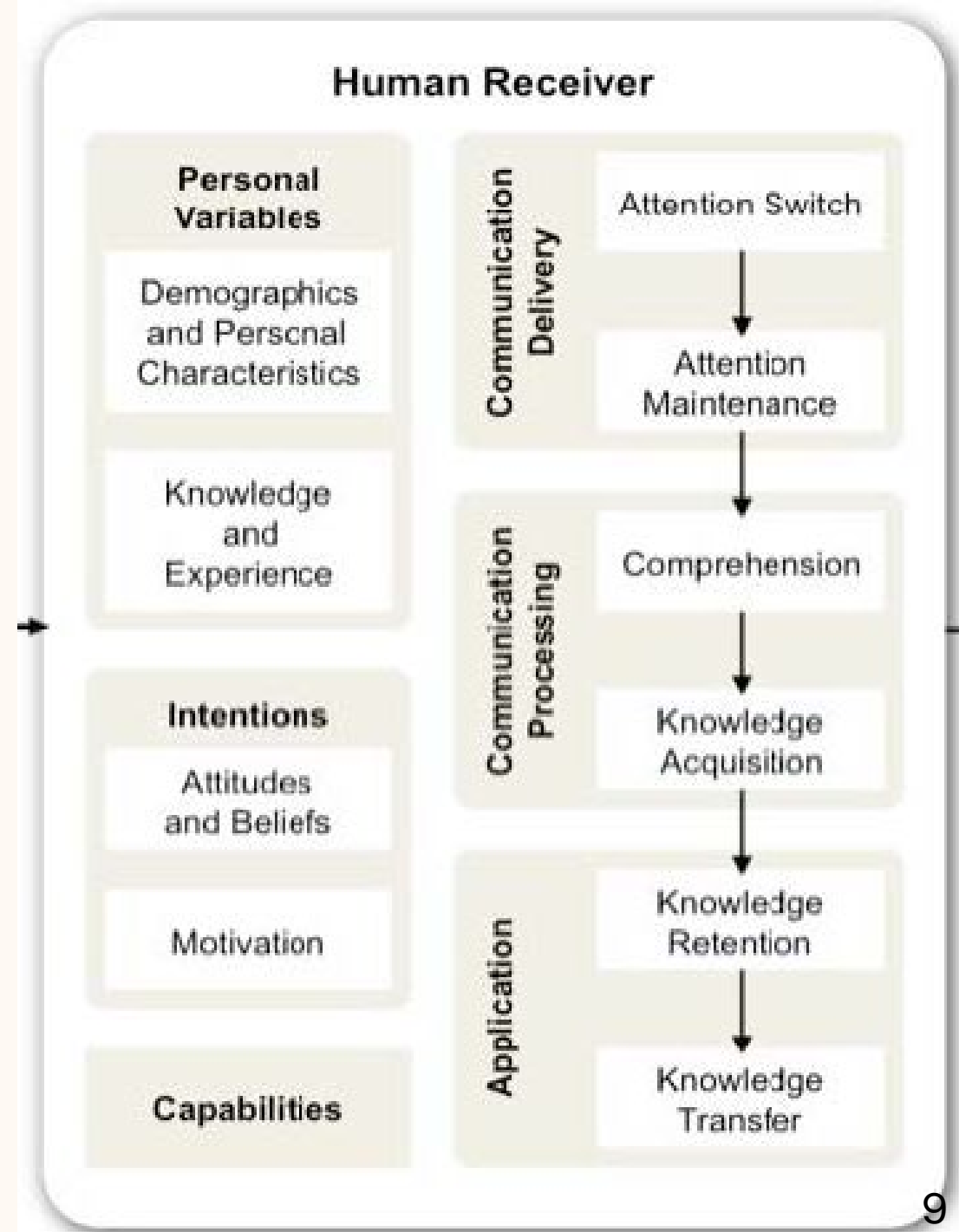
- Ask participants to answer a set of pre-defined questions.
- Pros:
 - gather data from a large number of people quickly
 - can determine how prevalent an issue or concern is
 - close-ended questions are easy to analyze
- Cons:
 - can only gather data you know about
 - careful planning is required before running a questionnaire
 - open-ended questions can take a lot of time to analyze and require careful setup

Questionnaires can be used at various points in the design process

- Understanding people
 - Understand the target population
 - Incorrect mental models
- Testing a theory
 - Are my assumptions correct?
 - Do people think that $A=B$?
- Testing a prototype design
 - How do people interpret my interface?
- Testing the final design
 - How are people actually using it?
 - What do people think after they use it?

What do you want to know?

- Attitudes
 - Do you like X?
 - Would using X work?
- Behaviors
 - How often do you use X?
 - Do you regularly do X?
- Knowledge
 - What is the best definition of X?
- Expectations
 - If the webpage did X what would you expect to happen?
- Capabilities
 - What is the result of adding 20 and 30?



Common survey elements

- Single and multiple choice checkboxes
- Matching
 - Rank the following from 1 to 5
- Rating scales
 - Likert Scales
 - 3, 5, 7 points scales
 - Semantic scales
- Open ended responses

OPEN ENDED

- Where does this URL go? What does it do?

Easier to write, harder to analyze

Harder to write, easier to analyze

CLOSE-ENDED

If you clicked on the link above, what web page would open?

- WWW3's main page
- National Geographic's main page
- World News's main page
- I will be taken to one of the sites above, but not their main page
- I will be taken to a website not listed above
- Other _____

Response Anchors
 Psychologists have been working for quite some time to determine the least biased way to present a set of answers.

On the right are a set of response anchors that are known to work well.

Likert-Type Scale Response Anchors

Citation:

Vagias, Wade M. (2006). *Likert-type scale response anchors*. Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management. Clemson University.

Level of Acceptability

- 1 – Totally unacceptable
- 2 – Unacceptable
- 3 – Slightly unacceptable
- 4 – Neutral
- 5 – Slightly acceptable
- 6 – Acceptable
- 7 – Perfectly Acceptable

Level of Appropriateness

- 1 – Absolutely inappropriate
- 2 – Inappropriate

My beliefs

- 1 – Very untrue of what I believe
- 2 – Untrue of what I believe
- 3 – Somewhat untrue of what I believe
- 4 – Neutral
- 5 – Somewhat true of what I believe
- 6 – True of what I believe
- 7 – Very true of what I believe

Priority:

Level of Support/Opposition

- 1 – Strongly oppose
- 2 – Somewhat oppose
- 3 – neutral
- 4 – Somewhat favor
- 5 – Strongly favor

Level of Probability

- 1 – Not probable
- 2 – Somewhat improbable
- 3 – Neutral
- 4 – Somewhat probable

Level of Acceptability

- 1 – Totally unacceptable
- 2 – Unacceptable
- 3 – Slightly unacceptable
- 4 – Neutral
- 5 – Slightly acceptable
- 6 – Acceptable
- 7 – Perfectly Acceptable

• 3 – Sometimes but

Affect on X

Frequency

Fill in the blank
type question

Q2: What is your age? _____

Typical multiple
choice
question

Q8: What is the highest level of education you have achieved?

- High school or less
- Some College
- Bachelor's Degree
- Master's Degree
- Doctorate Degree

Scale where
multiple
questions are
meant to be
summed
together

Q12: To what extent do you agree or disagree with each of the following statement

Please select one answer per row

	Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree
I often ask others for help with the computer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Others often ask me for help with the computer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Likert scale
question using a
pre-defined
anchor

Q13: In terms of your Internet skills, do you consider yourself to be:

- Not at all skilled
- Not very skilled
- Fairly skilled
- Very skilled
- Expert

System Usability Scale Questionnaire

**Strongly
Disagree**

**Strongly
Agree**

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

1	2	3	4	5
---	---	---	---	---

2. I found the product unnecessarily complex.

1	2	3	4	5
---	---	---	---	---

3. I thought the product was easy to use.

1	2	3	4	5
---	---	---	---	---

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

1	2	3	4	5
---	---	---	---	---

5. I found the various functions in the product were well integrated.

1	2	3	4	5
---	---	---	---	---

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

1	2	3	4	5
---	---	---	---	---

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

1	2	3	4	5
---	---	---	---	---

8. I found the product very awkward to use.

1	2	3	4	5
---	---	---	---	---

9. I felt very confident using the product.

1	2	3	4	5
---	---	---	---	---

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

1	2	3	4	5
---	---	---	---	---

System Usability Scale

#	Question	N/A	μ	σ
A1	<i>I apply software updates as soon as my computer prompts me.</i>	5 (1.0%)	3.20	1.221
A2	<i>I am happy to use an older version of a program, as long as it meets my needs.</i>	5 (1.0%)	^r 1.99	1.000
A3	<i>Whenever I step away from my computer, I lock the screen.</i>	5 (1.0%)	2.50	1.306
A4	<i>Others can access my smartphone or tablet without needing a PIN or passcode.</i>	21 (4.4%)	^r 3.34	1.545
A5	<i>When I discover a computer security problem at work, I'm likely to promptly report it to my employer.</i>	64 (13.4%)	4.08	0.995
A6	<i>It's important to use a WiFi password to prevent unauthorized people from using my home network.</i>	11 (2.3%)	4.66	0.690
A7	<i>I frequently click links in email messages to see what they are, regardless of who sent the message.</i>	5 (1.0%)	^r 4.51	0.922
A8	<i>It's important to run anti-virus software on my computer.</i>	7 (1.5%)	4.35	0.941
A9	<i>When browsing websites, I frequently mouseover links to see where they go, before clicking them.</i>	4 (0.8%)	4.13	0.977
A10	<i>When using public WiFi, I visit the same websites that I would visit when using the Internet at home.</i>	20 (4.2%)	^r 2.93	1.266
A11	<i>I usually do not pay attention to where I'm downloading software from.</i>	2 (0.4%)	^r 4.38	0.900
A12	<i>I frequently backup my computer.</i>	5 (1.0%)	3.07	1.165
A13	<i>I frequently visit websites even when my web browser warns me against it.</i>	8 (1.7%)	^r 3.98	1.028
A14	<i>I circumvent my employer's computer usage policies when they prevent me from completing a task.</i>	86 (18.0%)	^r 3.54	1.184
A15	<i>I am careful to never share confidential documents stored on my home or work computers.</i>	15 (3.1%)	4.36	0.757
A16	<i>Frequently checking the access control settings on social networking websites isn't worth the time it takes.</i>	18 (3.8%)	^r 3.56	1.165
A17	<i>I always write down my passwords to help me remember them.</i>	6 (1.3%)	^r 3.60	1.313
A18	<i>Creating strong passwords is not usually worth the effort.</i>	6 (1.3%)	^r 4.05	1.047
A19	<i>I frequently check my financial accounts for fraudulent charges.</i>	10 (2.1%)	4.11	0.914
A20	<i>If I receive a suspicious email from a company that I do business with, I'll phone the company to make sure the email is accurate.</i>	22 (4.8%)	3.52	1.236
A21	<i>I never give out passwords over the phone.</i>	7 (1.5%)	4.53	0.787
A22	<i>I frequently purchase things that I see advertised in unsolicited emails.</i>	4 (8.8%)	^r 4.51	0.840
A23	<i>I tend to ignore computer security stories in the news because they don't impact me.</i>	4 (8.8%)	^r 3.83	1.050
A24	<i>I use encryption software to secure files or email messages.</i>	10 (2.1%)	2.74	1.225
A25	<i>Once I create a password, I tend to never change it.</i>	5 (1.0%)	^r 3.30	1.182
A26	<i>I try to create a unique password for every account I have.</i>	5 (1.0%)	3.21	1.284
A27	<i>Rather than logging out of websites, I usually just navigate elsewhere or close the window when I'm done.</i>	7 (1.5%)	^r 3.06	1.299
A28	<i>I always make sure that I'm at a secure website (e.g., SSL, "https://", a lock icon) when transmitting information online.</i>	4 (0.8%)	3.80	1.173
A29	<i>I frequently use privacy software, "private browsing" or "incognito" mode when I'm online.</i>	6 (1.3%)	3.17	1.247
A30	<i>I frequently let others use my computing devices (e.g., smartphone, tablet, laptop).</i>	3 (0.7%)	^r 3.79	1.172

Table 1. Initial set of security questions evaluated on a 5-point Likert scale (from “strongly disagree” to “strongly agree”) by 479 participants. Depicted are the questions, the rate of “N/A” responses, and the average responses and standard deviations after recoding negatively-phrased questions (^r).

Scaling the Security Wall: Developing a Security Behavior Intentions Scale (SeBIS)

Planning a survey

Don't panic! This is not a statistics class.

COULD BE ON THE EXAM

- Independent and dependent variables
- Correlation vs causation
- Between vs within subject design
- Study question design

WILL NOT BE ON THE EXAM

- Statistical test names
 - T-test, ANOVA, etc.
- When to use different tests
 - Chi Sq should be used with categorical dependent and independent variables
- P-values, distributions, confidence intervals or other outcomes from tests

Topics Outline

- **Descriptive questions vs testing a question**
- Correlation vs causation
- Dependent vs independent variables
- Between and within subjects testing
- Numeric vs categorical data

Planning a survey

- Surveys normally answer **multiple research questions**. With each research question tied to one or more survey questions.
- **Descriptive** – learn something about the whole population.
 - How many people have heard of the term “phishing”?
 - What words do people use to describe cookie tracking?
- **Testing for correlation or causation** – show that two things are related or one thing causes the other thing.
 - If someone has been trained on phishing in the past, are they better at differentiating phishing emails?
 - We have three training options, each user goes through one training, which training causes people to identify phishing emails the best?

Descriptive Statistics

- **Descriptive Questions** – learn something about the **whole population**.
 - How many people have heard of the term “phishing”?
 - What words do people use to describe cookie tracking?
- **Descriptive Numeric** – fancy term for all the basic measures of numeric data: **Mean, median, mode, standard deviation**
 - What % of consumers are worried about privacy?
 - What % of people know the difference between behavioral advertising and cookies?
 - On average, how long does it take to decide if an email is phishing or not?
- **Descriptive Qualitative** – use data to learn about a whole population
 - What is the most common reason people avoid using ATMs?
 - Why do some people choose to not have a Google account?

Testing for correlation or causation

- Testing **for correlation or causation** – show that two things are related, or one thing causes the other thing.
 - If someone has been trained on phishing in the past, are they better at differentiating phishing emails?
 - We have three training options, each user goes through one training, which training **causes** people to identify phishing emails the best?
- These tests require more complex statistics, such as:
 - T-test
 - ANOVA
 - Linear Models
 - CHI Squared

Topics Outline

- Descriptive questions vs testing a question
- **Correlation vs causation**
- Dependent vs independent variables
- Between and within subjects testing
- Numeric vs categorical data

Correlation vs. Causation

- Correlation

- Two things tend to behave in a way that seems inter-related, where if one thing changes the other thing will also change in a related way.
- For example, if the price of rice goes up at the same time as the price for beans.

- Causation

- When one thing changes it causes the other thing to change.
- For example, when the weather gets cold more people wear coats.
Cold weather causes more people to wear coats.

Does consuming chocolate increase the number of Nobel Laureates?

This is a correlation, not necessarily a causation.

Chocolate Consumption, Cognitive Function, and Nobel Laureates
Franz H. Messerli, M.D.

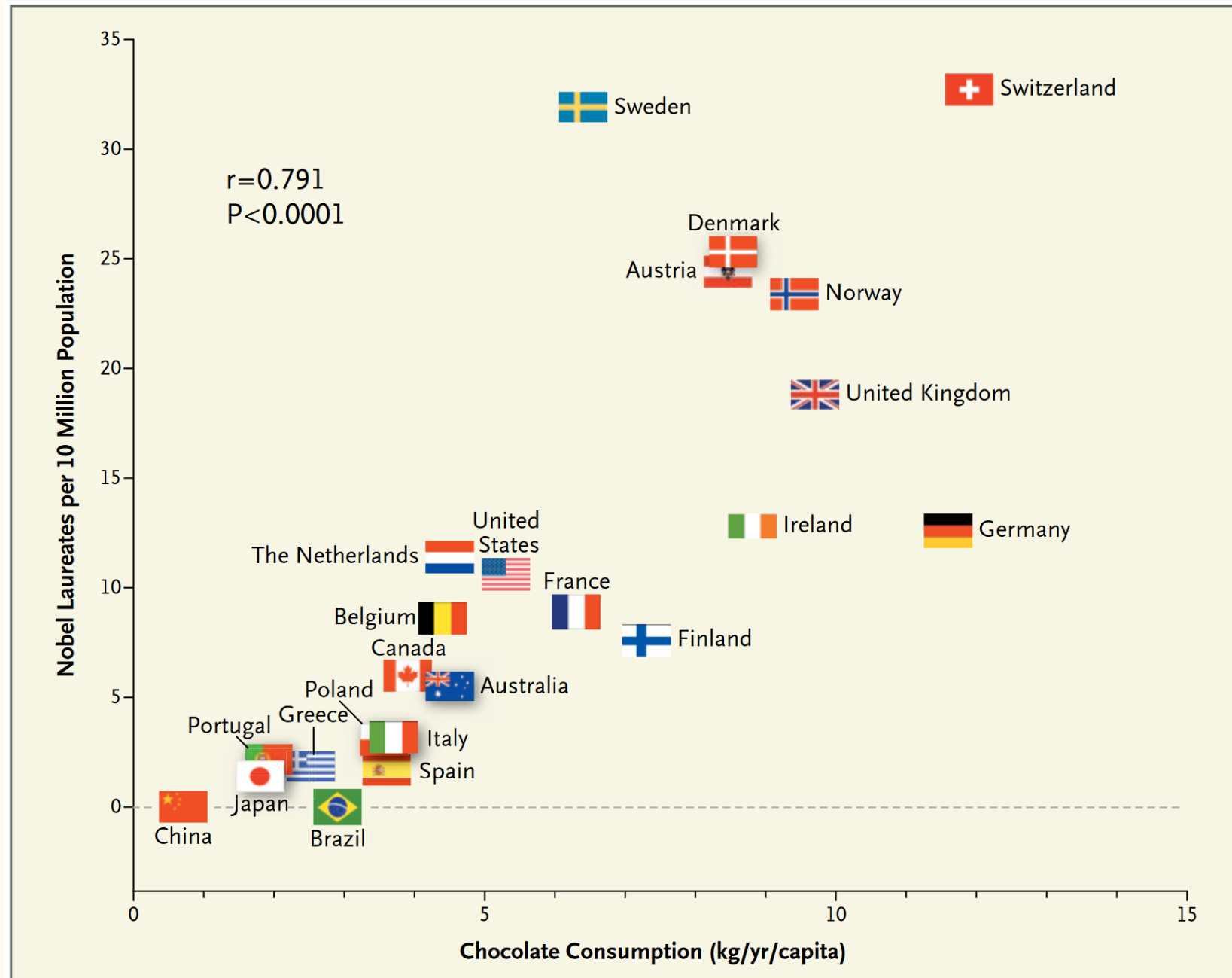
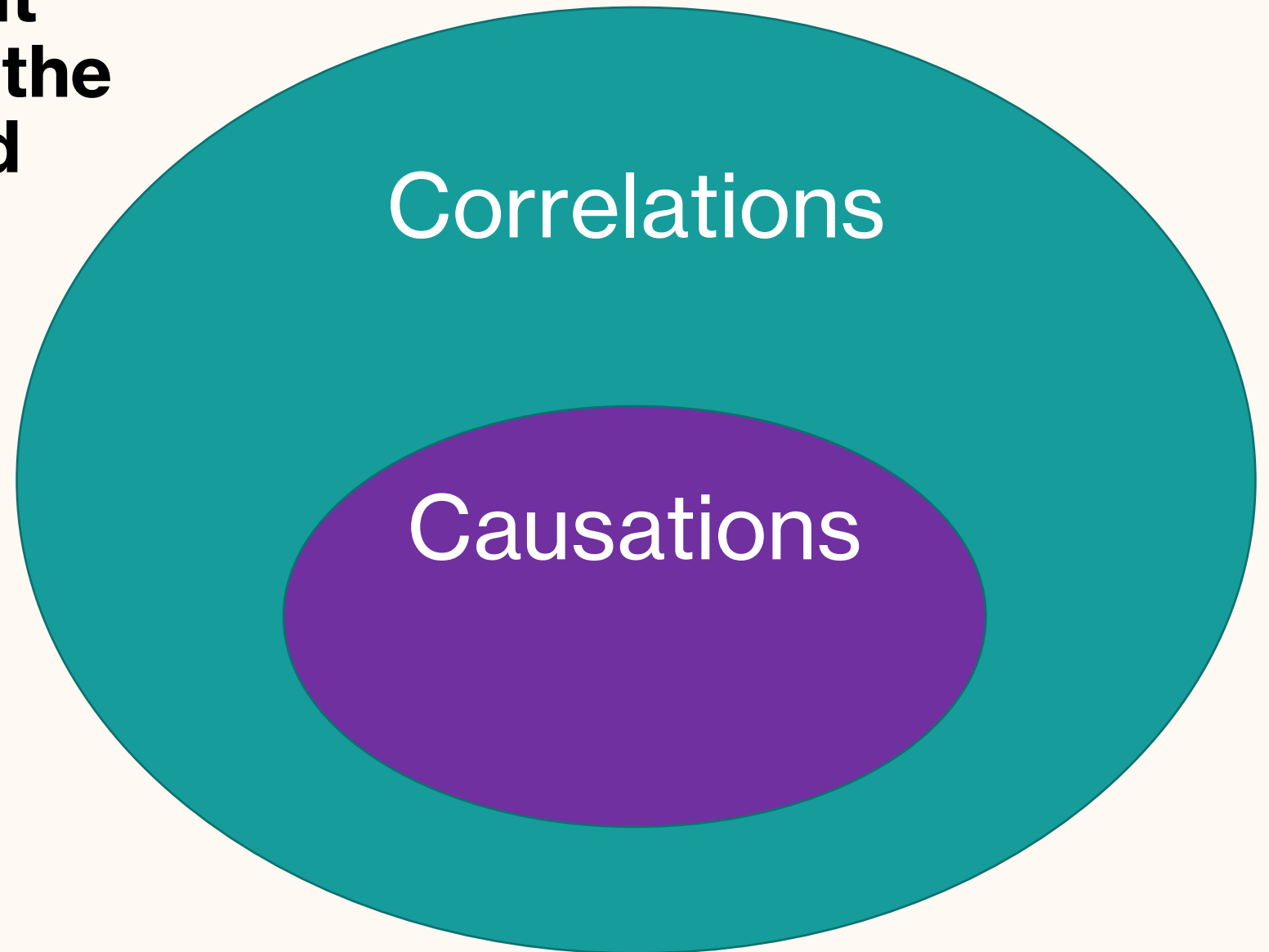


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

**Causations are
Correlations, but
not necessarily the
other way round**



**History +
CTR is a
correlation**

**How might
you test if
it is really
a
causation?**

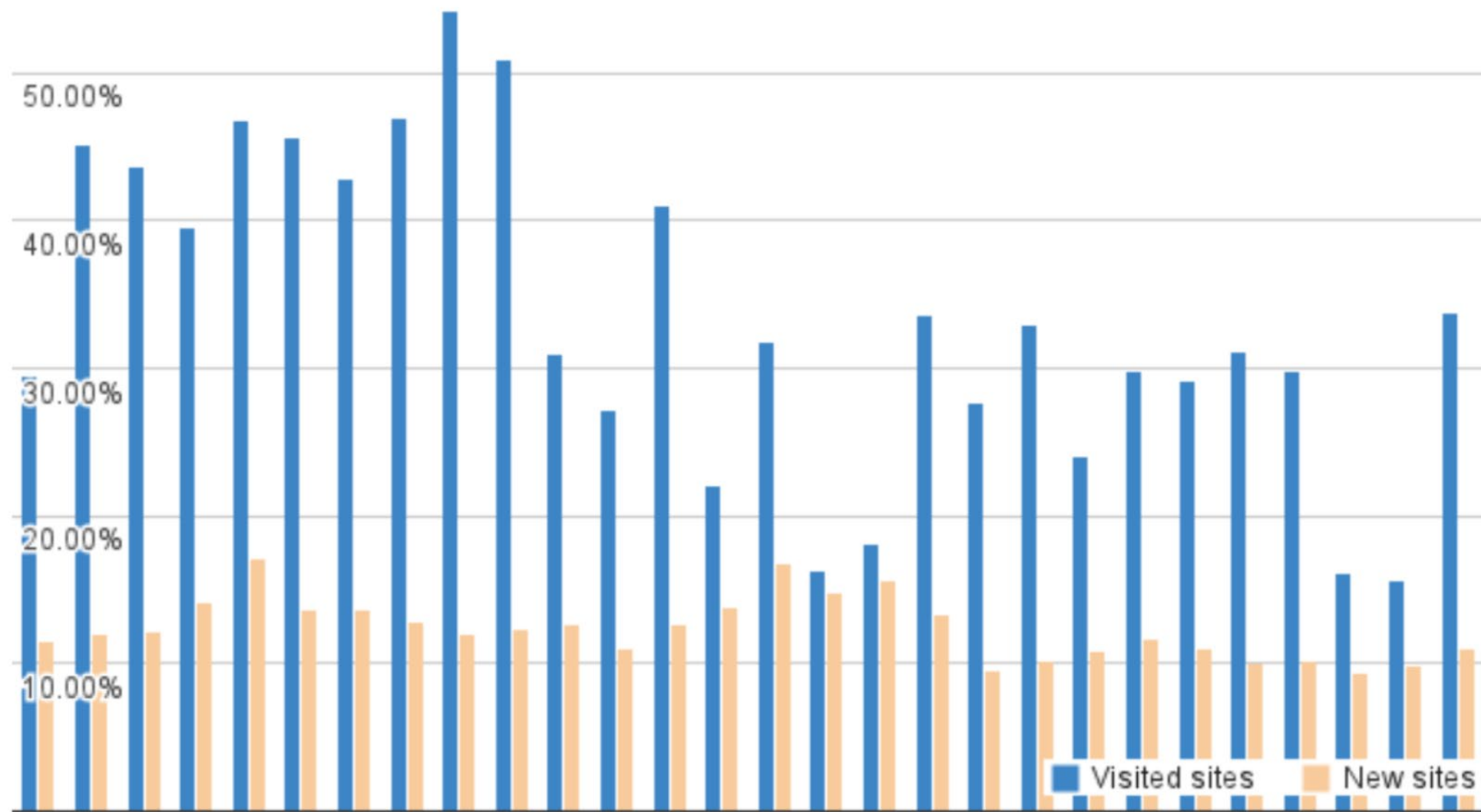


Figure 3: Daily CTR, separated by whether the website was already in the user's browsing history. For 28 days in January-February 2014.

Topics Outline

- Descriptive questions vs testing a question
- Correlation vs causation
- **Dependent vs independent variables**
- Between and within subjects testing
- Numeric vs categorical data

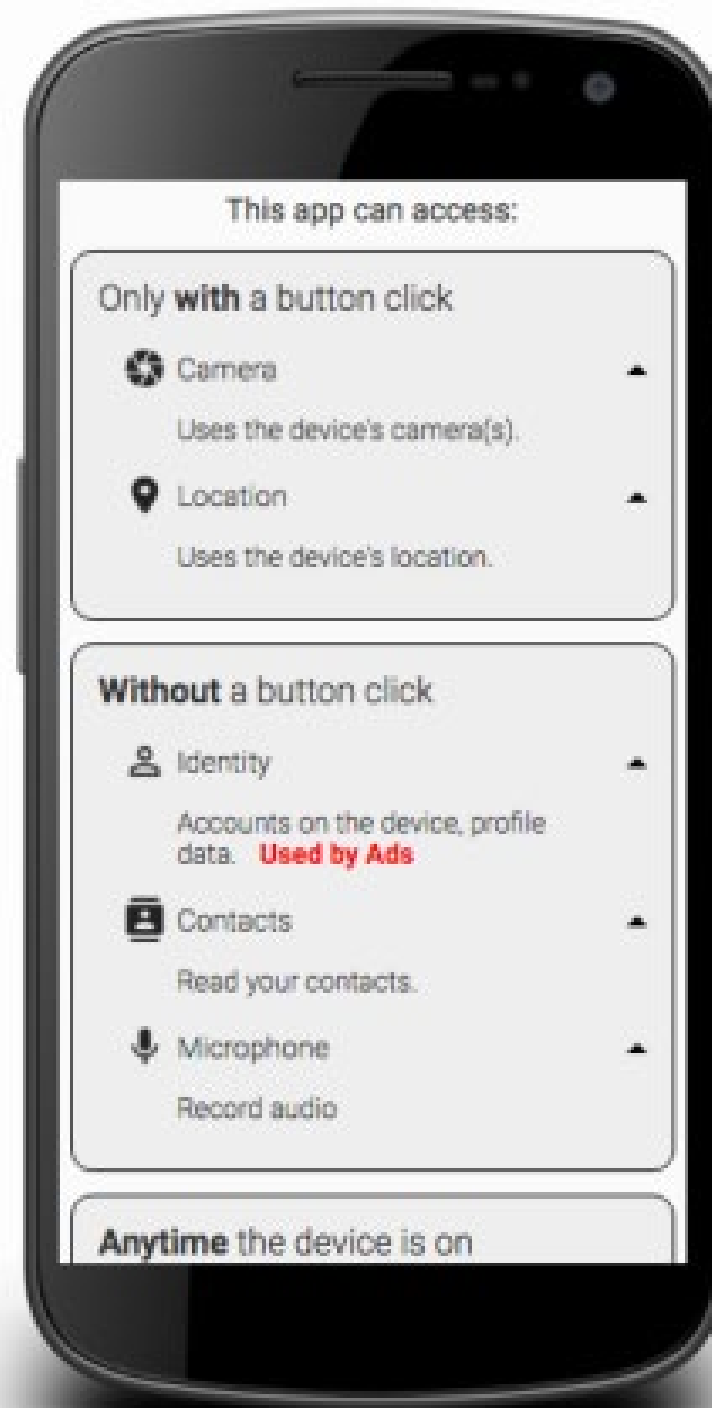
What are you going to measure?

- In statistics there are classically two types of measurements (variables): dependent and independent
- Dependent
 - Also known as the **outcome variable**
 - “Dependent” on the study
 - Measures the usability **goal**
- Independent
 - Anything **you are directly manipulating**
 - An element of the study which is under your control
 - A pre-existing feature of your participant

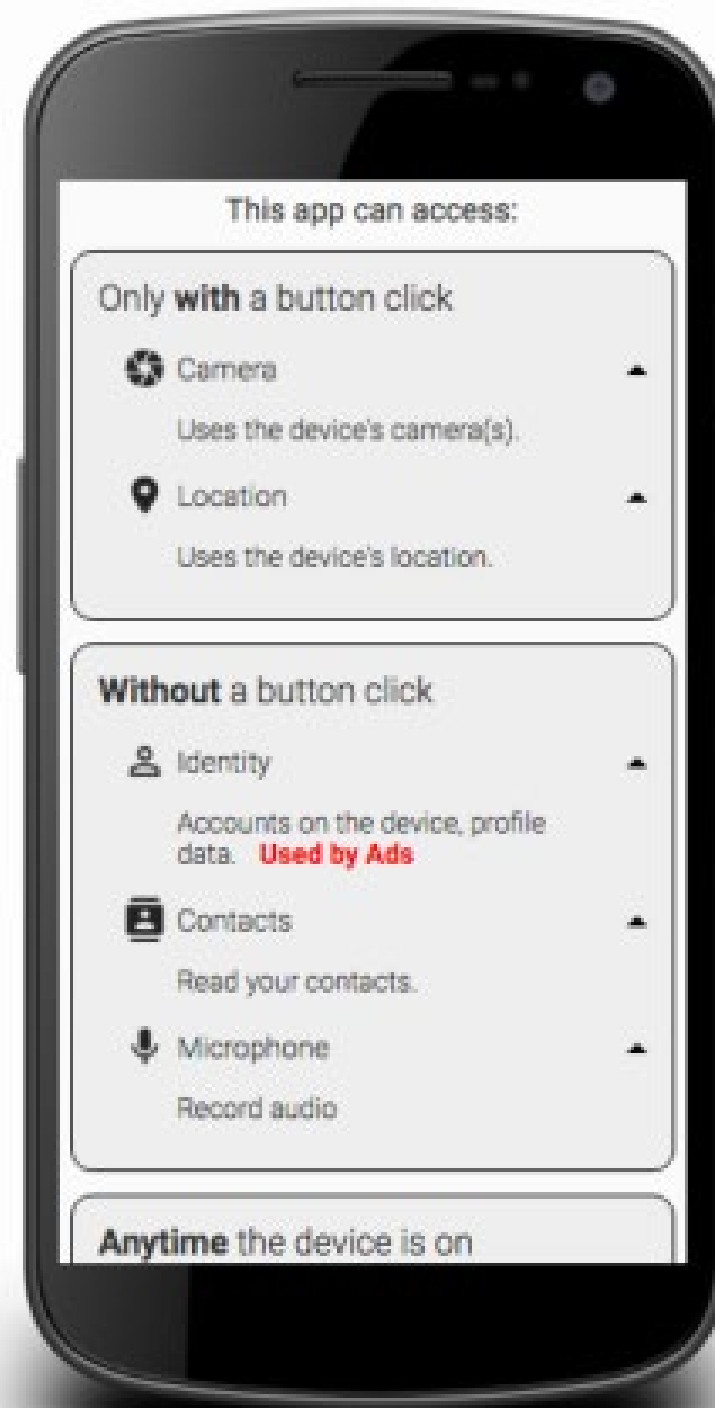
Some research questions:

- Can people differentiate between a subdomain and a domain when reading a URL?
- Does [a new system] help people differentiate between malicious URLs and safe ones?
- Can users use [a new password manager] faster and with less errors than [the old password manager]?
- Does knowing how an app will use its permissions impact app installation decisions?
- Using [website], can users successfully opt-out of cookie tracking without forming inaccurate mental models?

Lets use this study as an example



Research Question:
Can users reliably identify if an app can or cannot perform an action directly tied to a permission.





Awesome App

can access

- Location
Uses the device's location
- Camera
Uses the device's camera(s)



Awesome App

can access

Without a button click

- Microphone
Record audio
- Camera
Uses the device's camera(s).
- Location
Uses the device's location. **Used by Ads**

Dependent variable:
Count of the number of questions the participant answered correctly

g can this app do?

Independent variable:
Which of the two interfaces the participant was shown

Absolutely
Possible

- Charge purchases to your credit card at any time.
- Get your location.
- Allow ads to know your location.
- Load ads.
- Write on the SD card

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Variables that would make sense

- Research Question: Can users reliably identify if an app can or cannot perform an action directly tied to a permission?
- Dependent
 - Which permissions correctly/incorrectly read
 - Count of permissions correctly/incorrectly read
 - Time spent reading each permission screen
- Independent
 - Study group (which screen was shown)
 - If the permission was privacy sensitive or not
 - Order of the tasks
 - Time of day
 - Type of most used device (laptop, mobile, PC)
 - Demographics of the participants (gender, age, native language, ...)

Common dependent things to measure

- Number of dangerous errors made
- Time to complete task
- Percent of task completed
- Percent of task completed per unit of time
- Ratio of successes to failures
- Time spent in errors
- Percent or number of errors
- Percent or number of competitors better than it
- Frequency of help and documentation use

Topics Outline

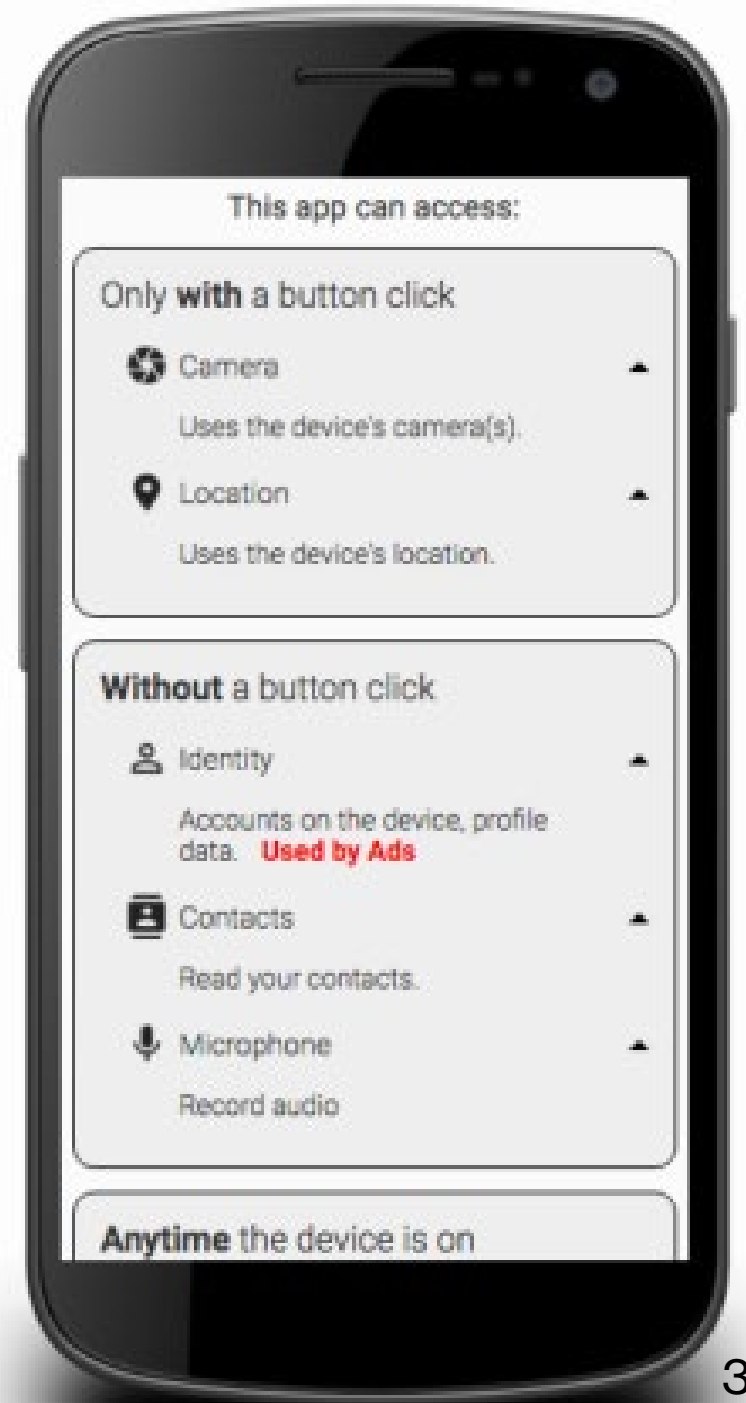
- Descriptive questions vs testing a question
- Correlation vs causation
- Dependent vs independent variables
- **Between and within subjects testing**
- Numeric vs categorical data

Between vs. Within subjects

- Between subjects
 - Your study only shows one interface to one person
 - You are measuring how well the people randomly assigned to the A interface did compared to the people randomly assigned to the B interface
 - **Lots of variability with this method**
- Within subjects
 - Your study shows all interfaces to all people
 - You are measuring the difference in how they do on the two interfaces
 - **Less variability (same person) but more learning effects and priming**

Study design

- RQ: Does [my new interface] enable people to accurately determine what permissions an app will use?
- A/B test between the existing and new interface
- Between subjects
- 10 Tasks shown in the same order to all participants
- Dependent variables
 - Accuracy on task
- Independent variables
 - Which interface (A or B)



Topics Outline

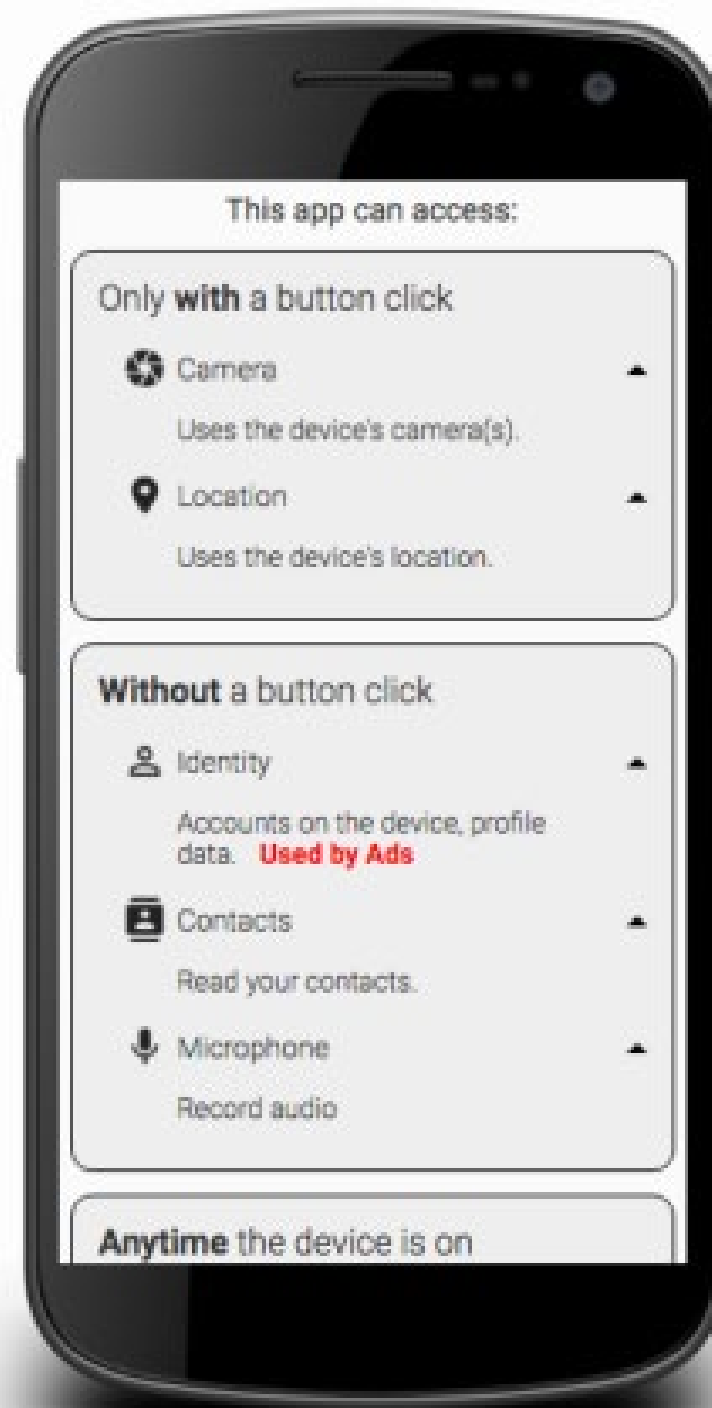
- Descriptive questions vs testing a question
- Correlation vs causation
- Dependent vs independent variables
- Between and within subjects testing
- **Numeric vs categorical data**

Types of data

- Numeric
 - **Continuous** – Any value on the range is possible including decimal (1-5)
 - **Discrete** – Only certain values on the range are possible (1,2,3,4,5)
 - **Interval** – Only certain values on the range are possible and each has equal distance from its neighboring values (strongly agree, agree, neutral, disagree, strongly disagree)
- Categorical
 - **Binary** – Only two possibilities (true, false)
 - **Ordinal** – The values have an ordering (slow, medium, fast)
 - **Nominal** – The values have no ordering (apple, pear, kiwi, banana)

Study design

- Accuracy on all tasks
 - Discrete
- Which interface
 - Categorical binary



Statistical tests

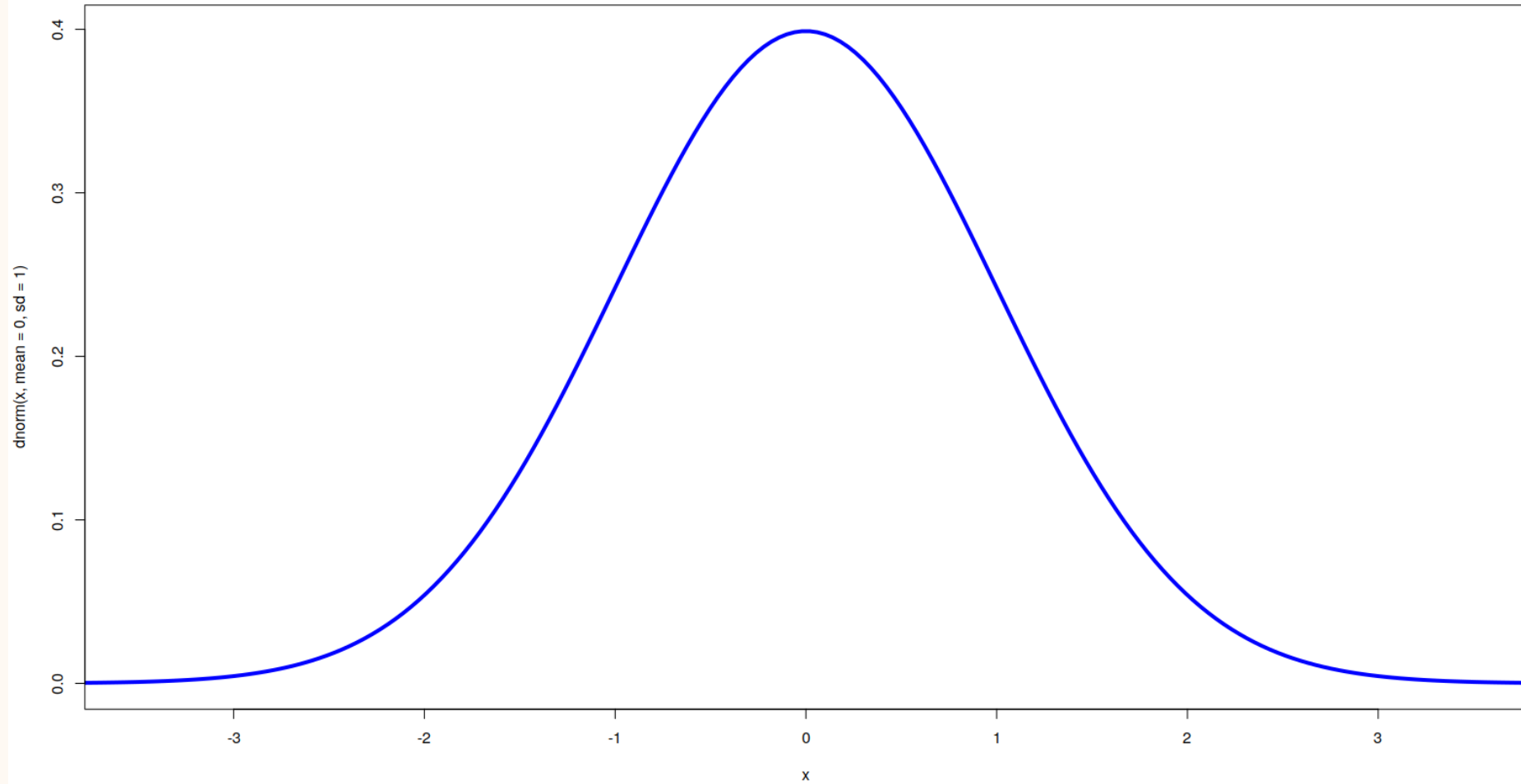
Comparing	Dependent	Independent	Parametric (Dependent variable is mostly normally distributed)	Non-parametric
The means of two independent groups	Continuous / scale	Categorical / nominal	Independent t-test	Mann-Whitney test
The means of 2 paired (matched) samples	Continuous / scale	Time variable (before/after)	Paired t-test	Wilcoxon signed rank test
The means of 3+ independent groups	Continuous / scale	Categorical / nominal	One-way ANOVA	Kruskal-Wallis test
3+ measurements on the same subject	Continuous / scale	Time variable	Repeated measures ANOVA	Friedman test
Relationship between 2 continuous variables	Continuous / scale	Continuous / scale	Pearson's Correlation Coefficient	Spearman's Correlation Coefficient
Predicting the	Continuous /	Any	Simple Linear	

**t-test: Test if two groups have the same mean
(average)**

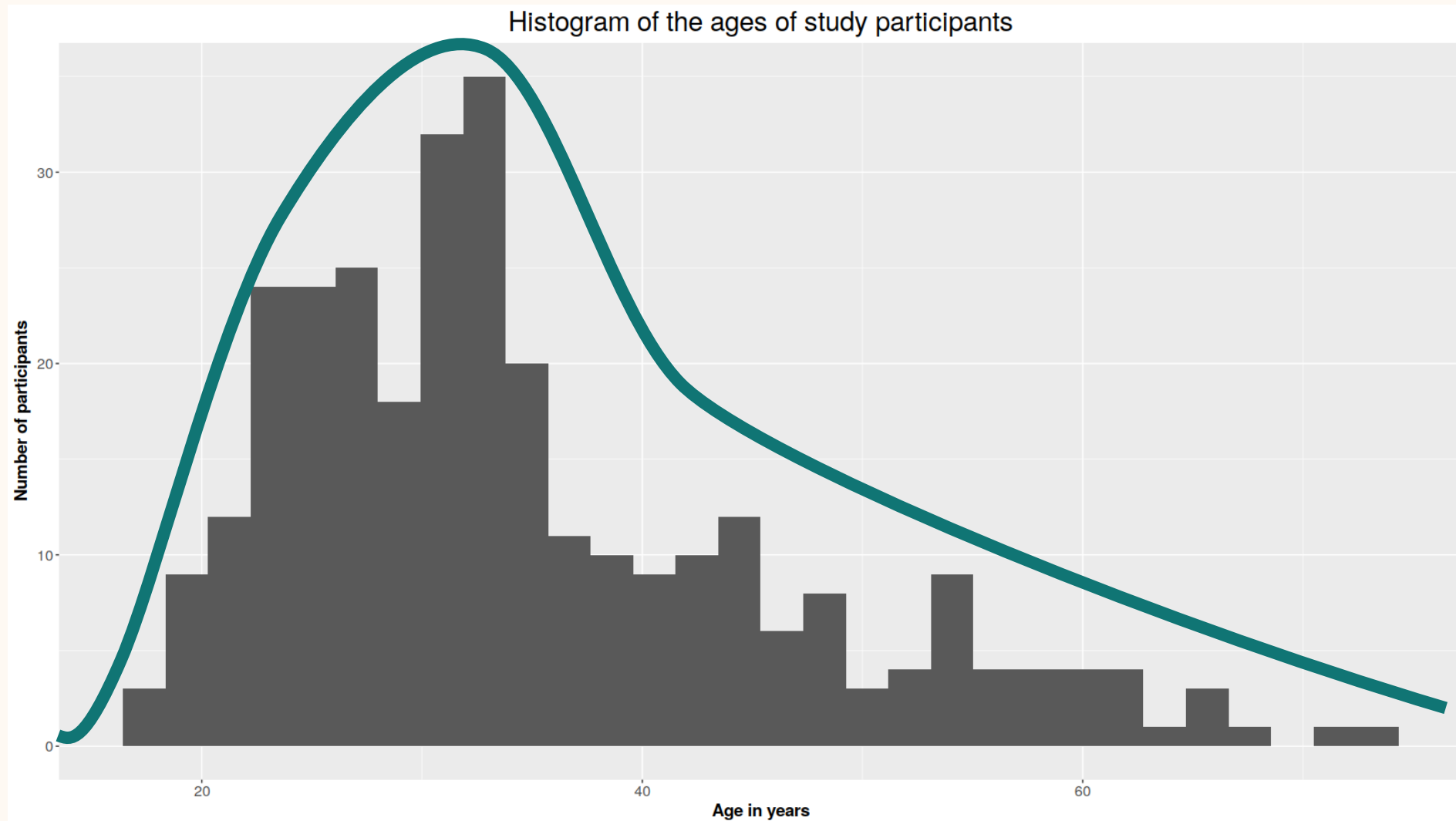
T-test requires:

- Independent variable: categorical binary
- Dependent variable: numeric (continuous or discrete)
- Data must be **normally distributed**

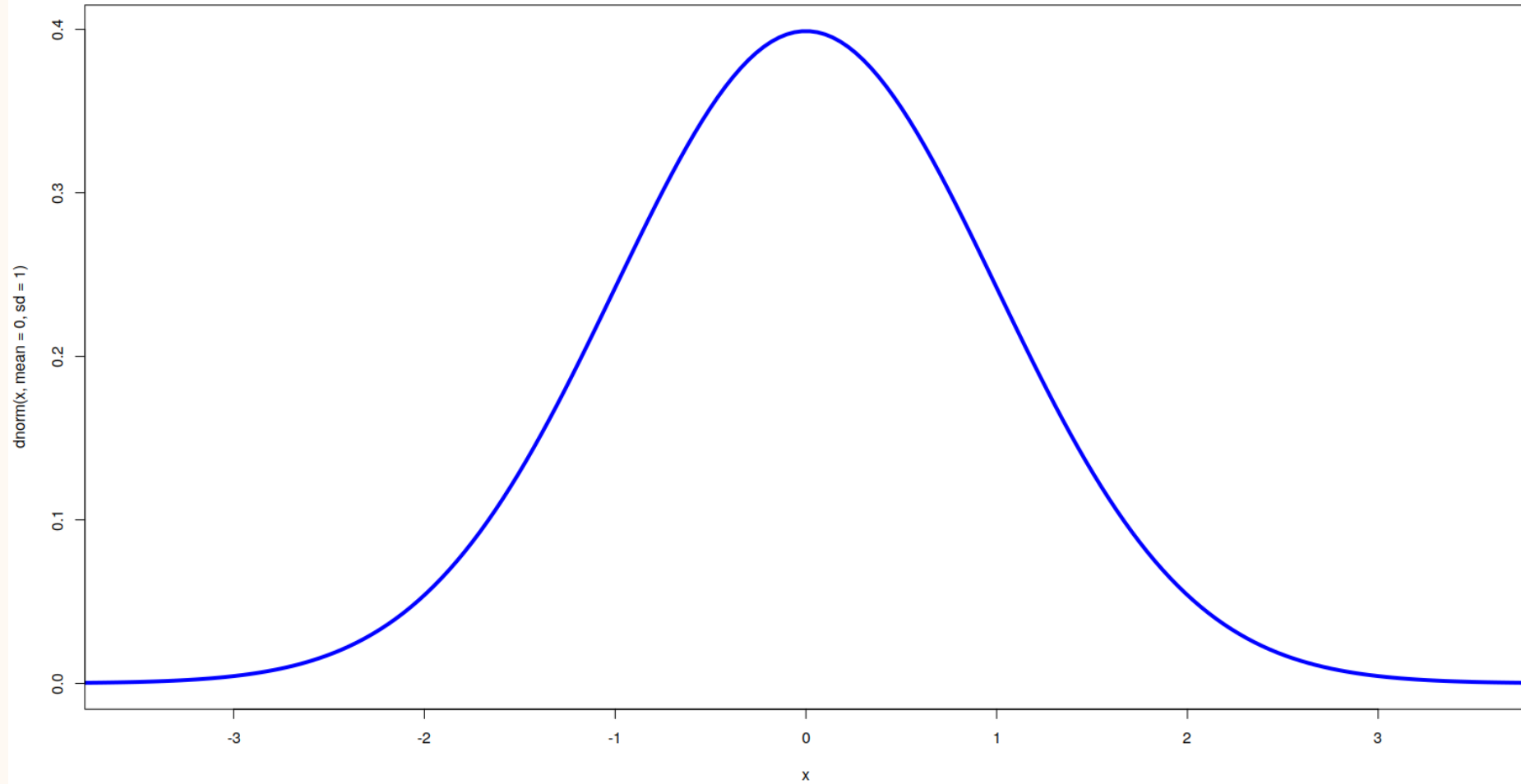
Normal distribution



Real data is messy

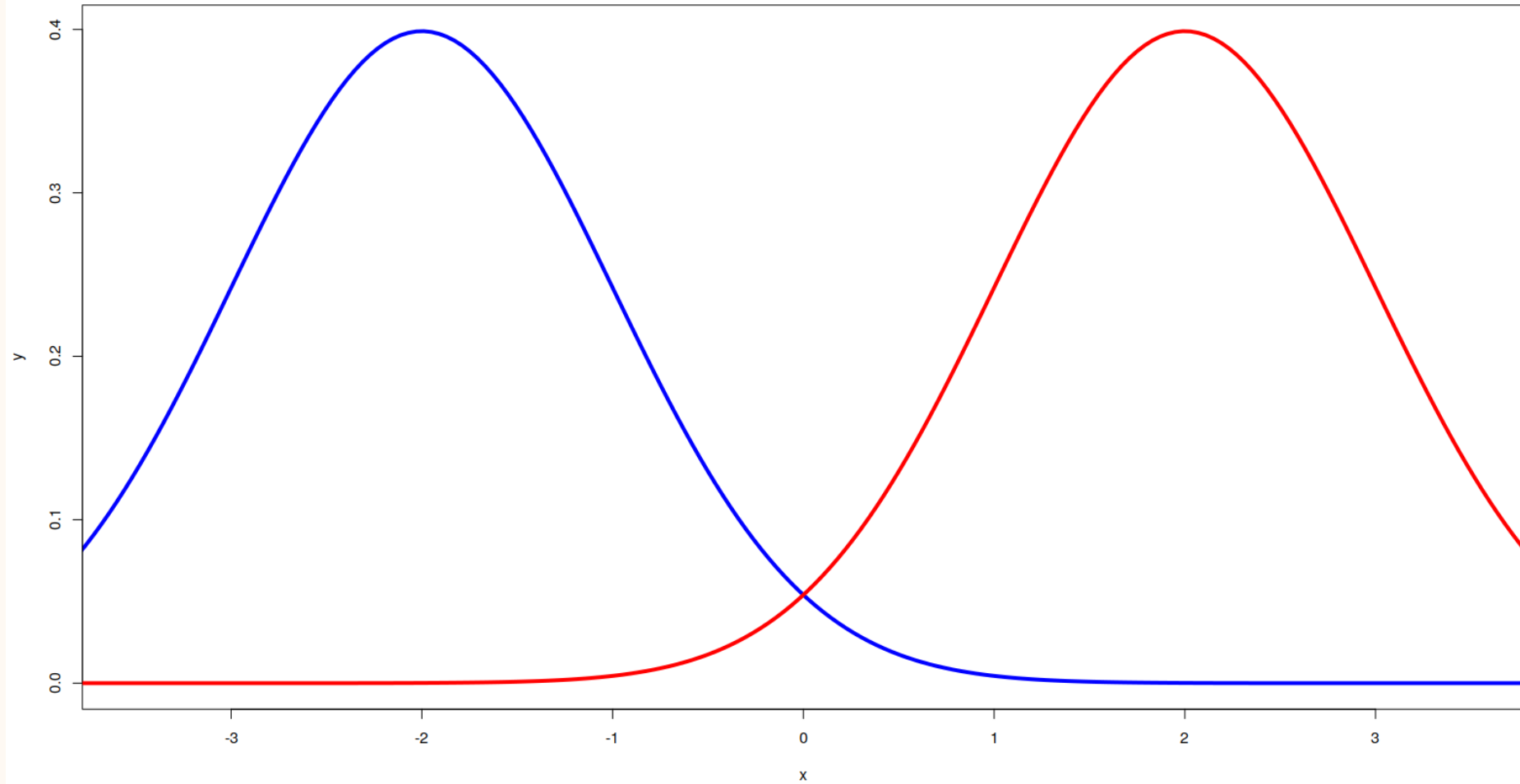


Normal distribution

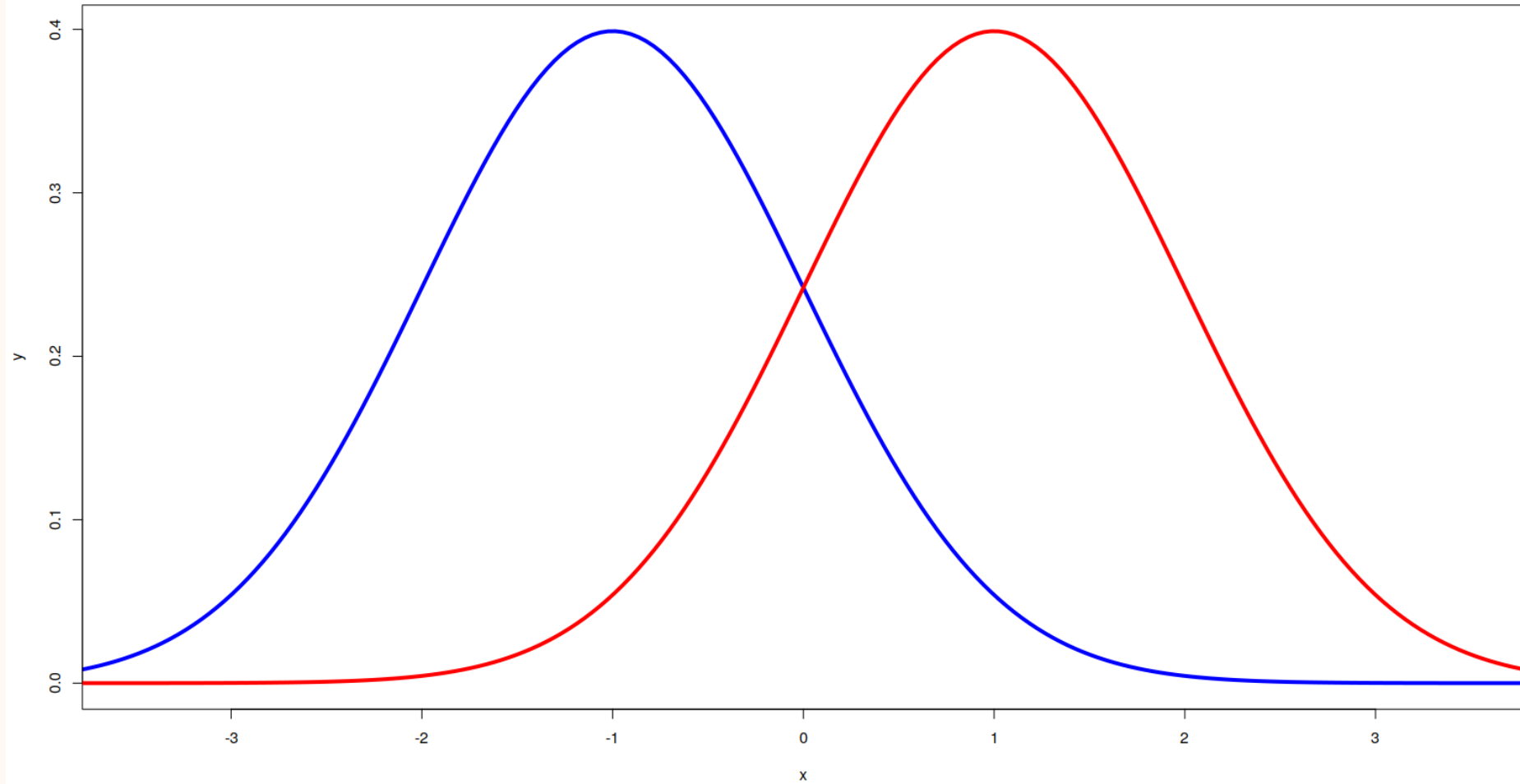


T-test: Do two populations have the same mean?

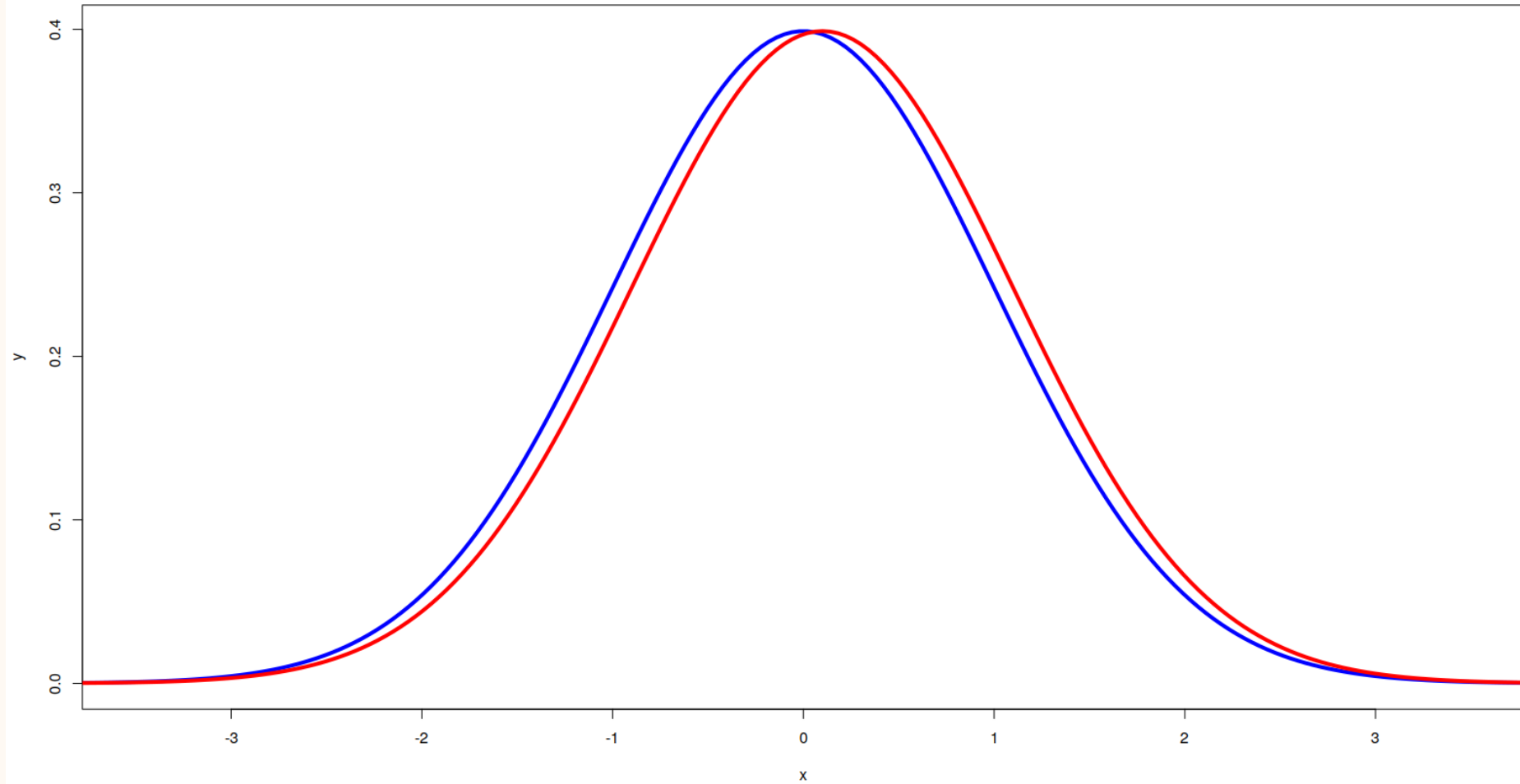
Different means



Maybe? different means



Likely not different means



I showed participants 4 code samples and asked them what the code would do. I then asked them how confident they were in their answer.

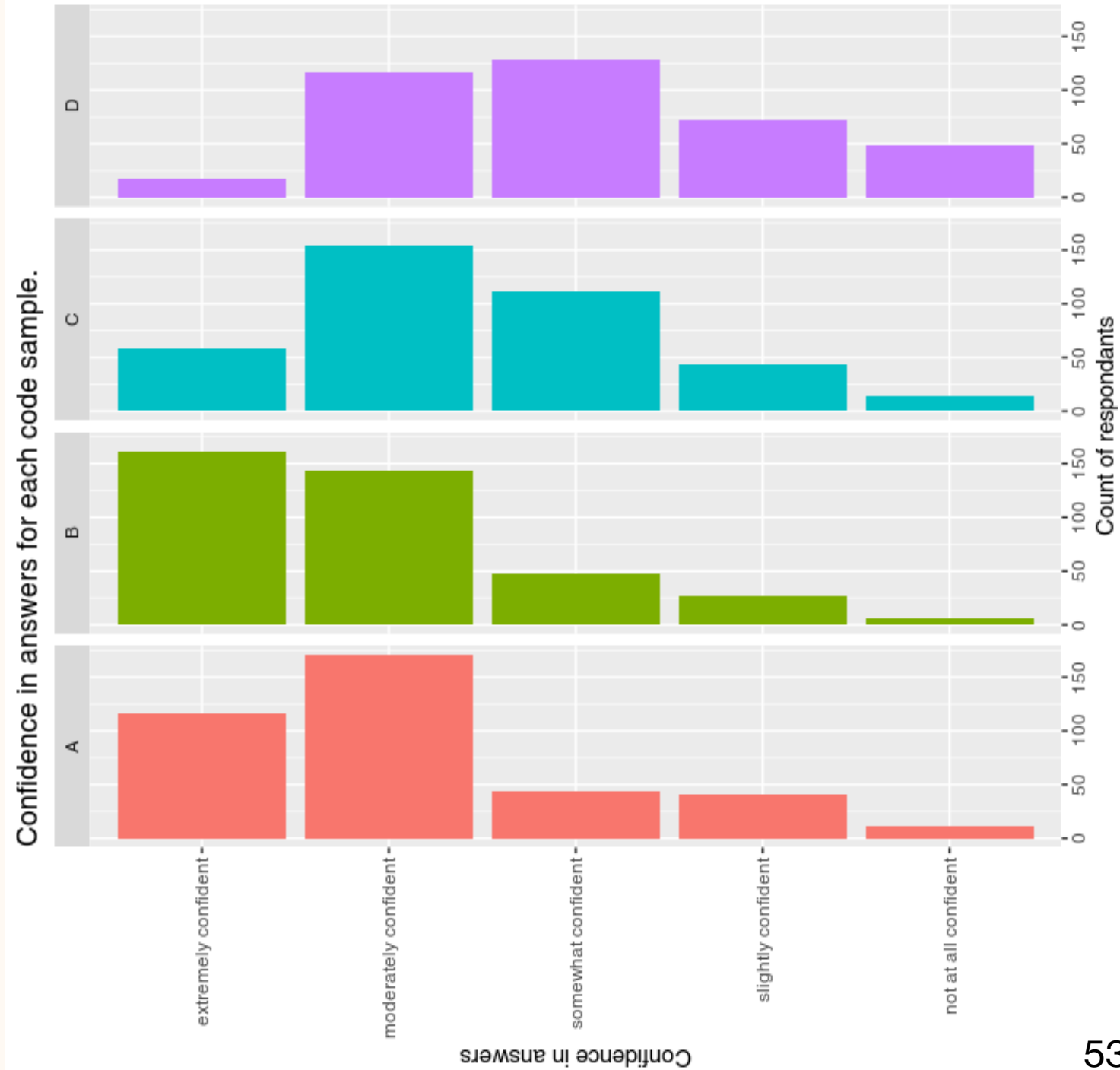
Research Question: Does the code sample shown impact confidence in their answer?

Research Question:
Does the code sample shown impact confidence in their answer?

Within-subjects

Independent:
Which code sample shown

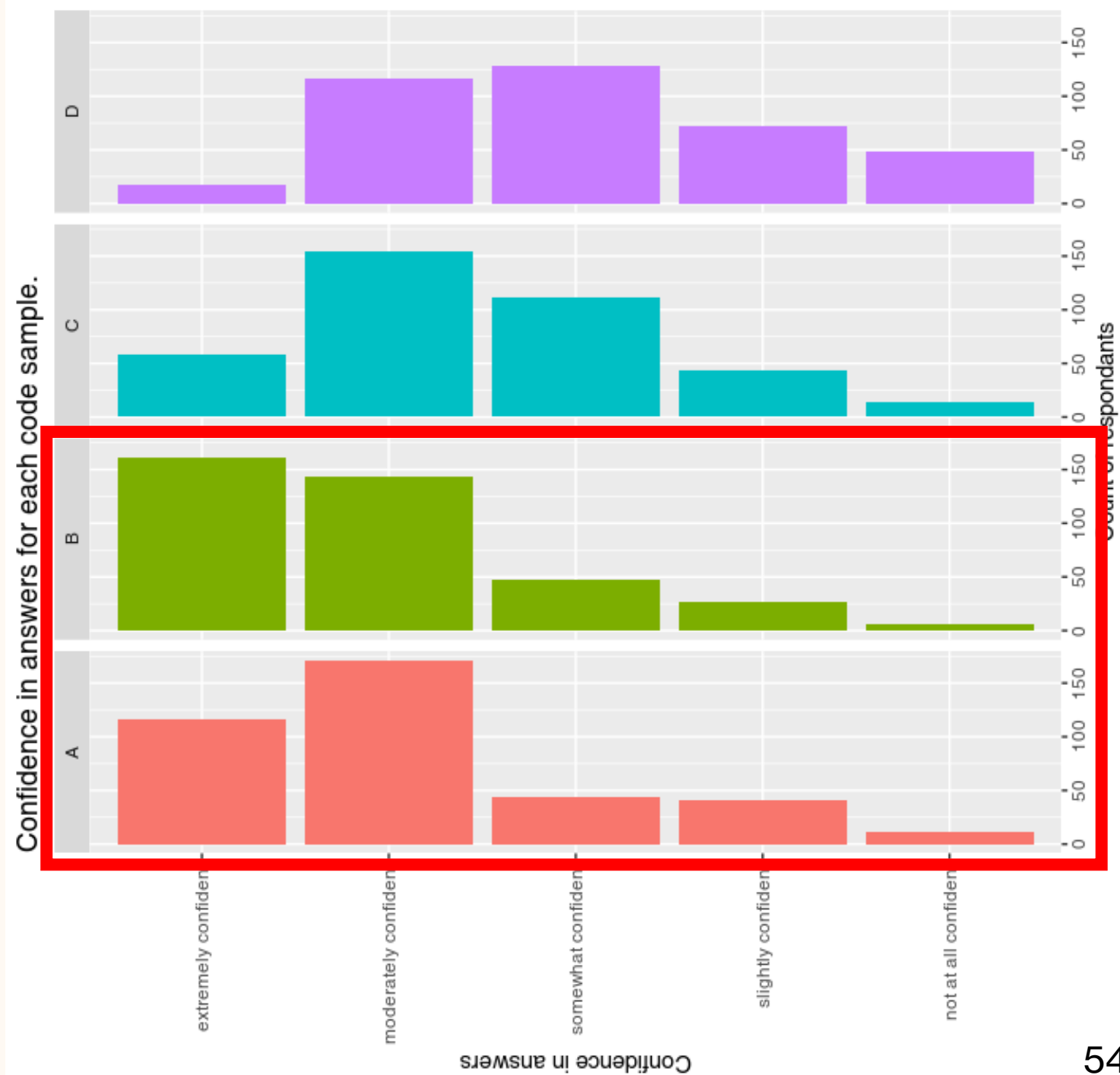
Dependent:
Confidence



Problem: My categorical variable (code sample) is not binary, there are 4 levels.

Solution: Run the t-test on each pair. So test A vs B, A vs C, C vs D.

Real solution: Use an ANOVA (not covered in this class)



Running the t-test

- This is a “**within** subjects” test where one person gave a confidence answer for **both** Code Sample A and Code Sample B
 - So we use a **Paired t-test**
- Create two arrays (or Excel columns) one with Code Sample A confidence, the other with Code Sample B confidence
- Two-sided (tailed)
 - For now, just do this. I don't have time to explain.
- Alpha of 0.05
 - p-value needs to be less than 0.05 to show that the two code samples produce different levels of confidence
 - Means that 5% of the time we will get the wrong answer from the statistical test

```
> t.test(a.confidence,b.confide
```

Paired t-test

```
data: a.confidence and b.confidence
```

```
t = -5.2699, df = 383, p-value = 2.285e-07
```

```
alternative hypothesis: true difference in means is not equal to 0
```

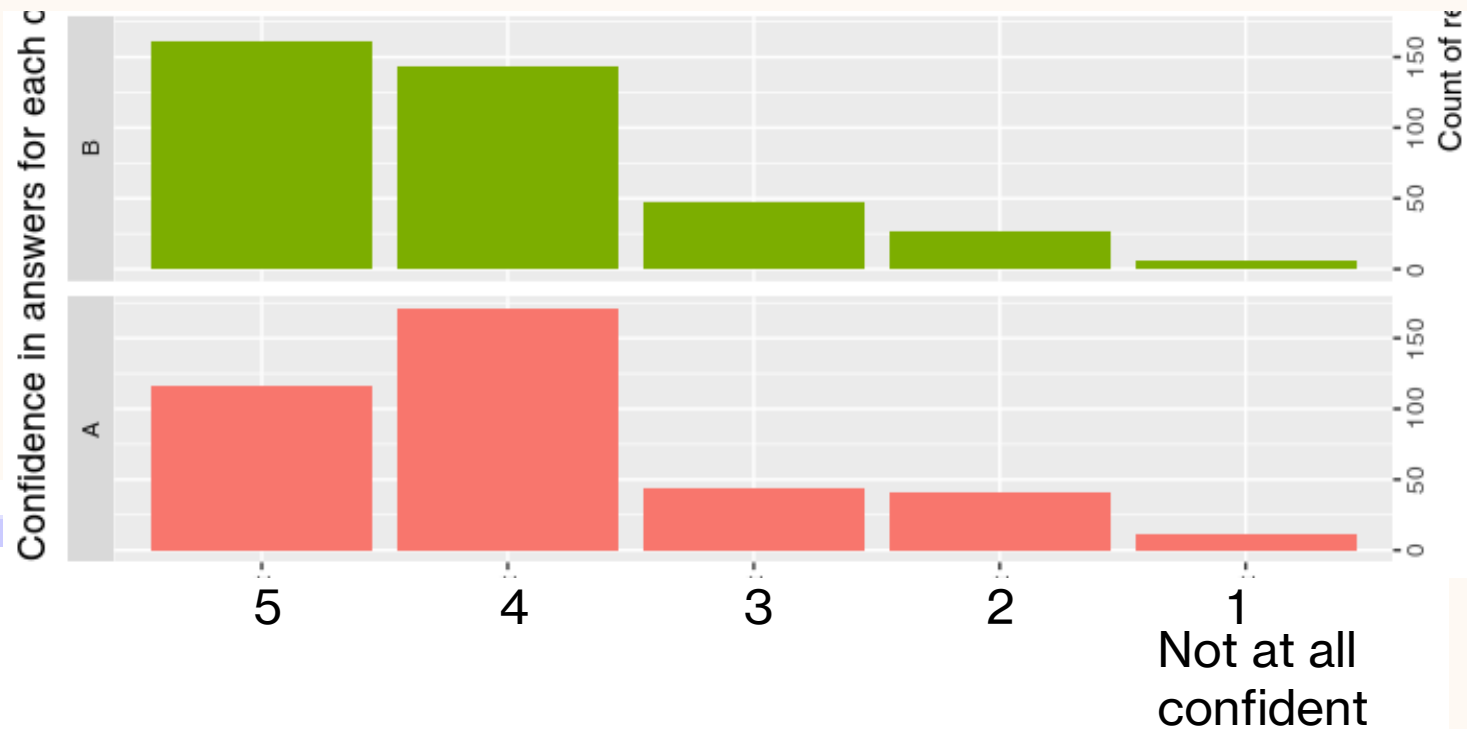
```
95 percent confidence interval:
```

```
-0.3218198 -0.1469302
```

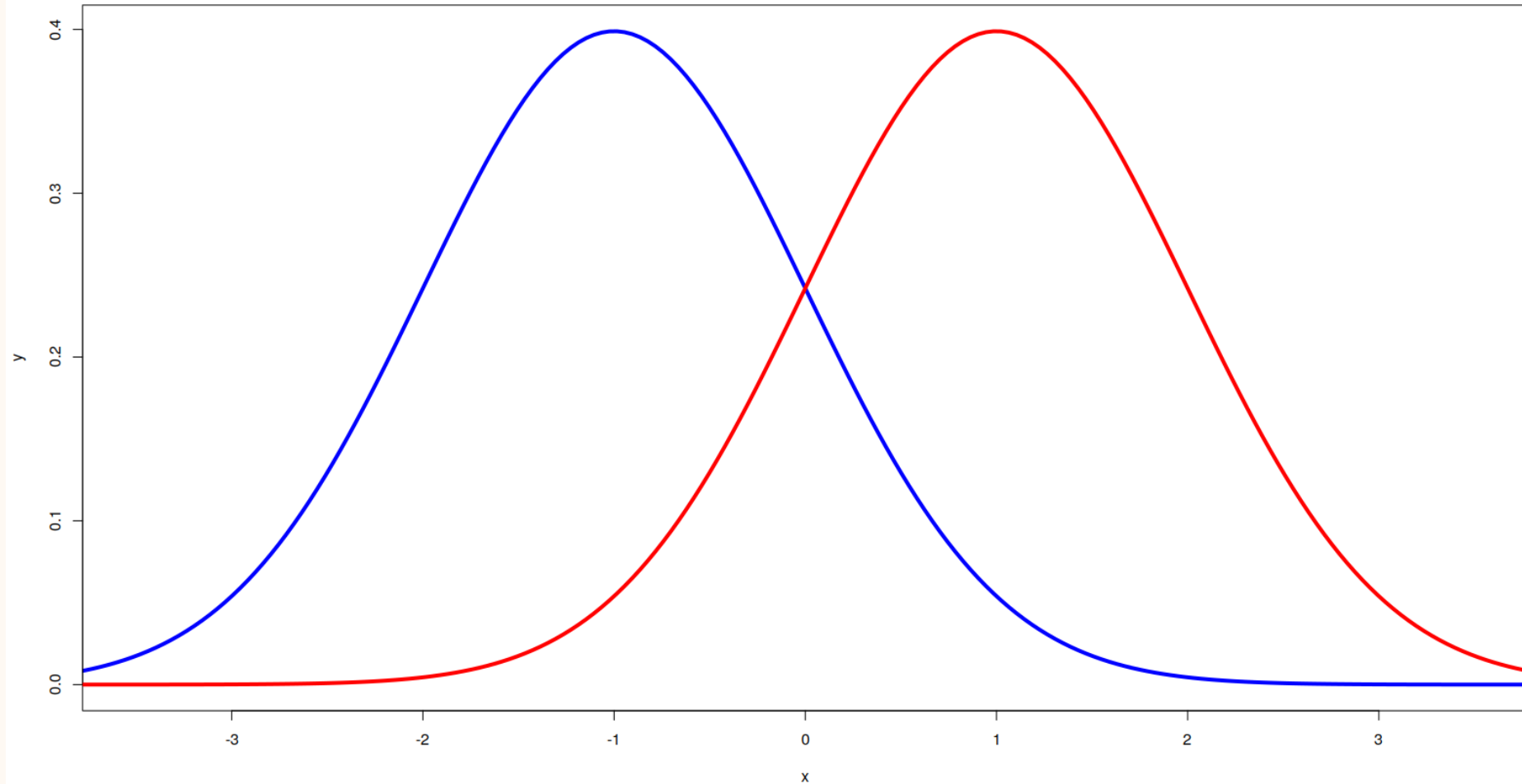
```
sample estimates:
```

```
mean of the differences
```

```
-0.234375
```



Different means, small difference



I ran a survey to learn about software update behaviors.

Research Question: Do women and men feel like they ask others for technical help with different frequency?

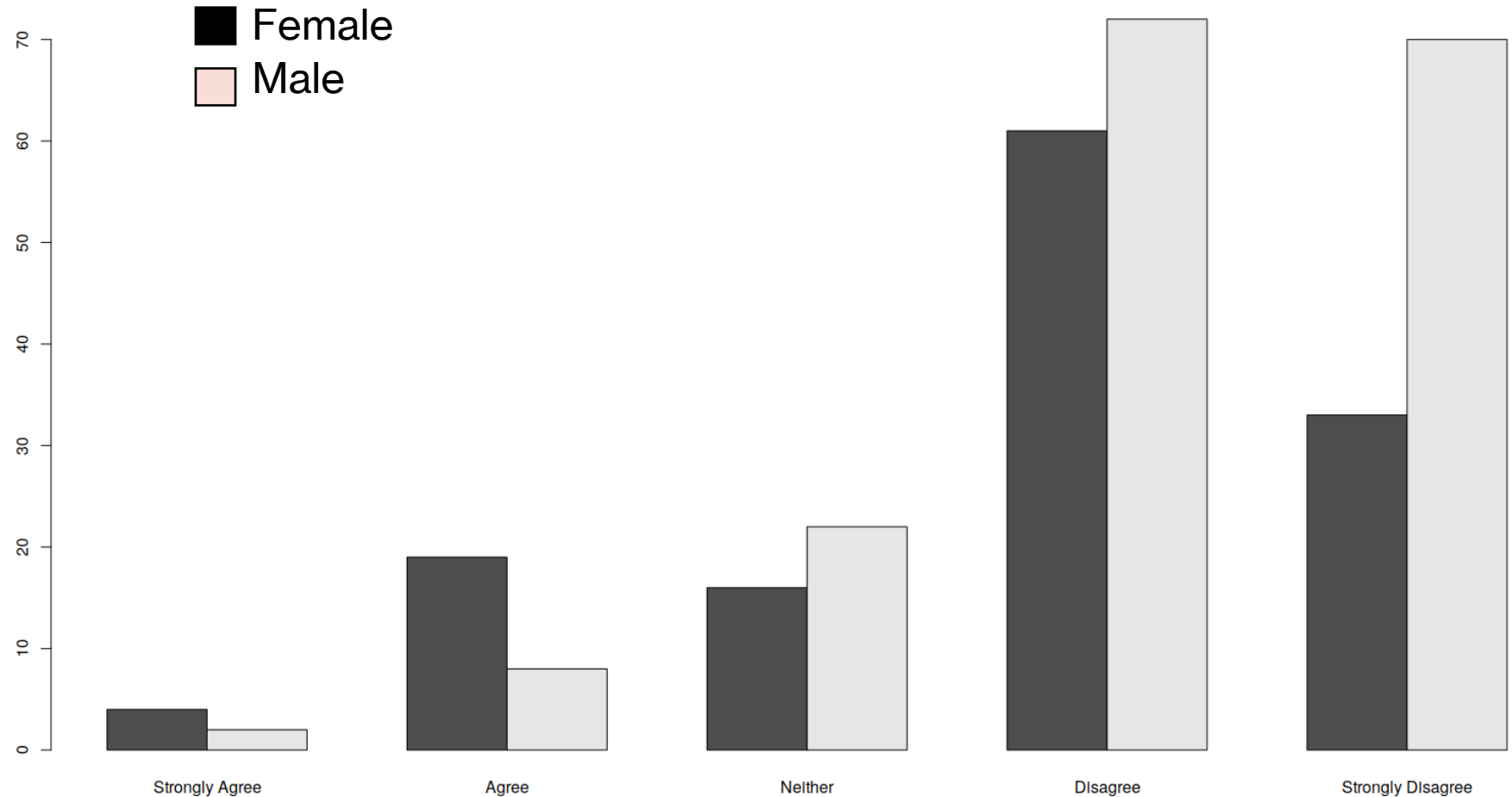
Research

Question: Do women and men feel like they ask others for help with different frequency?

Between-subjects

Independent:
Gender

Dependent:
Agreement



I often ask others for help with technical questions

Running the t-test

- This is a “**between** subjects” test where each person gave only one answer
 - So we use a **normal t-test** (not paired)
- Create two arrays one with women’s responses, one with men’s
- Two-sided (tailed)
 - For now, just do this. I don’t have time to explain.
- Alpha of 0.05
 - p-value needs to be less than 0.05 to show that the two genders produce different levels of confidence
 - This choice means that 5% of the time we will get the wrong answer from the statistical test

```
> t.test(as.numeric(d$i_ask_others_for_he
```

Welch Two Sample t-test

```
data: as.numeric(d$i_ask_others_for_help[d$gender == "Female"]) and [d$gender == "Male"]  
t = -3.4481, df = 253.99, p-value = 0.0006606
```

```
alternative hypothesis: true difference in means is not equal to 0
```

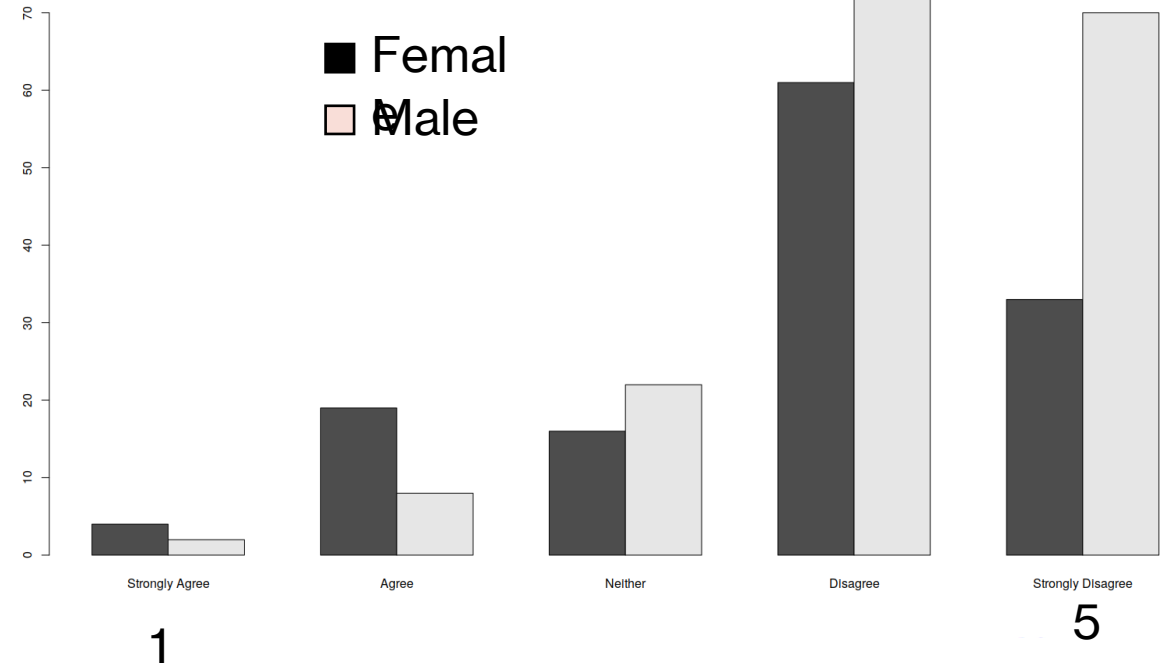
```
95 percent confidence interval:
```

```
-0.6245978 -0.1704934
```

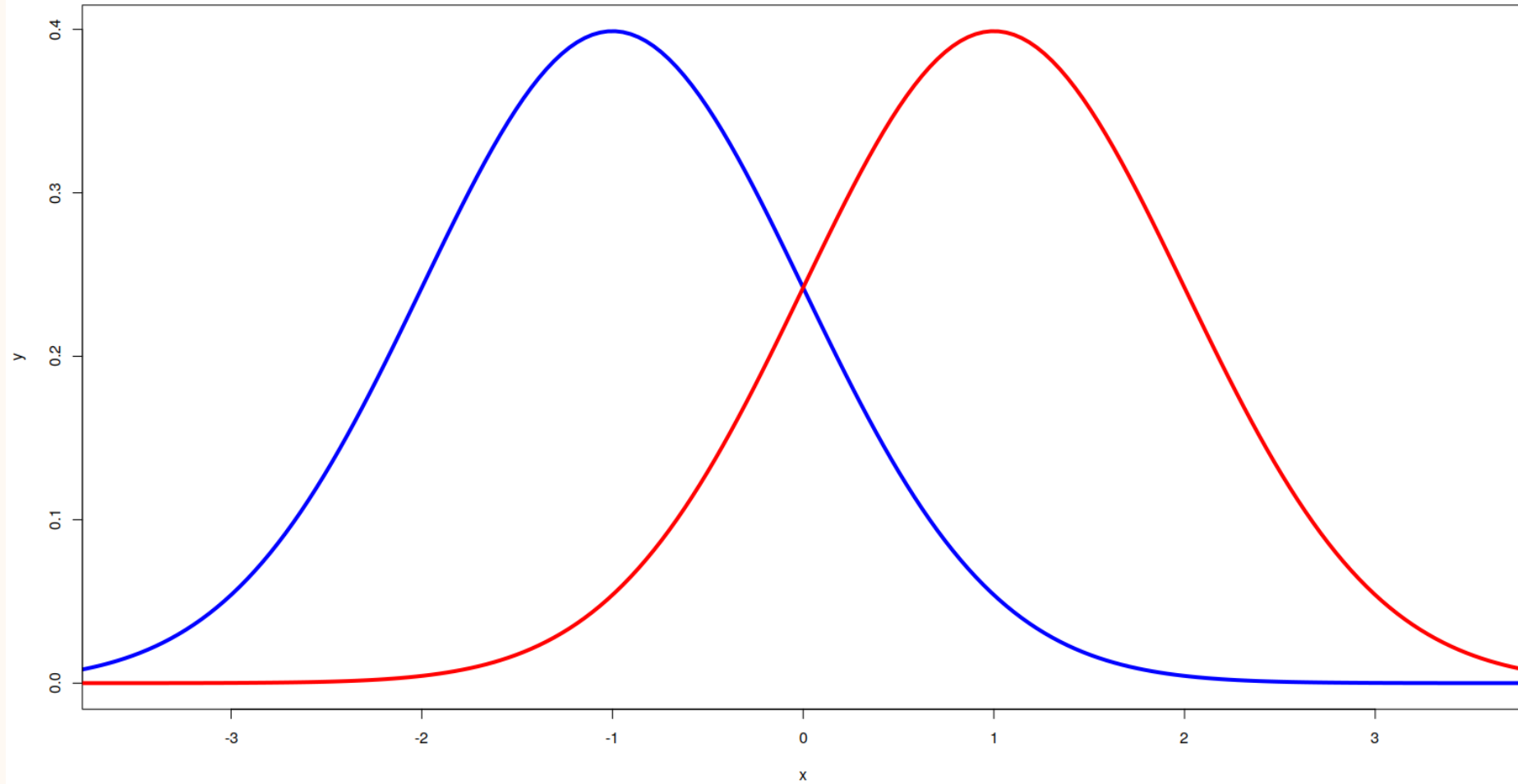
```
sample estimates:
```

```
mean of x mean of y
```

```
3.751880 4.149425
```



Maybe? different means



I asked participants to tell me a story about a prior software update.

Research Question: Are people who relate positive stories older or younger?

Research

Question: Are people who relate positive stories older or younger?

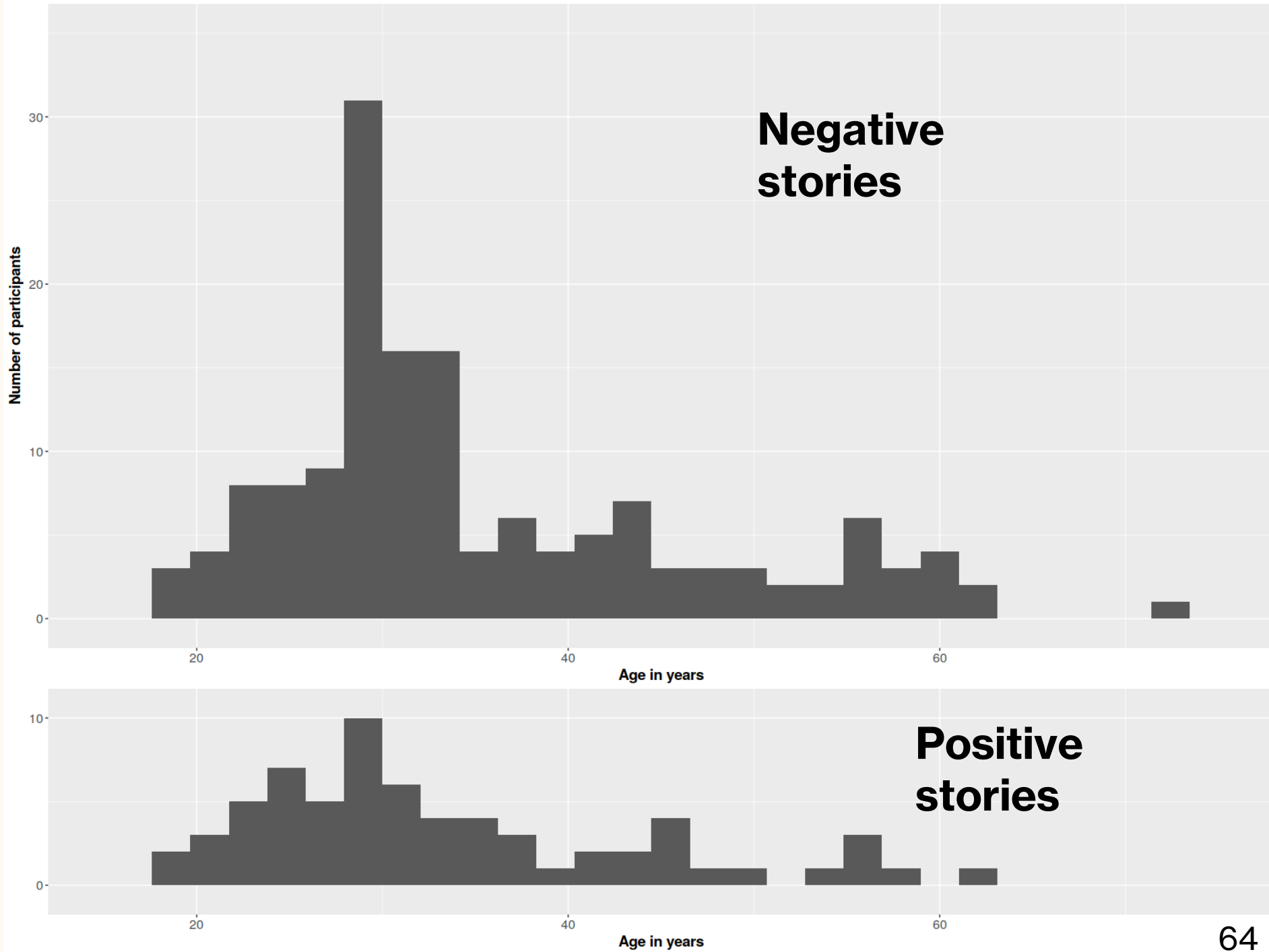
Between-subjects

Dependent:

- Age
- Numerical

Independent:

- Negative or Positive
- Binary




```
> t.test(s_neg$age, s_pos$age)
```

```
Welch Two Sample t-test
```

```
data: s_neg$age and s_pos$age
```

```
t = 0.75677, df = 123.07, p-value = 0.4506
```

```
alternative hypothesis: true difference in means is not equal to 0
```

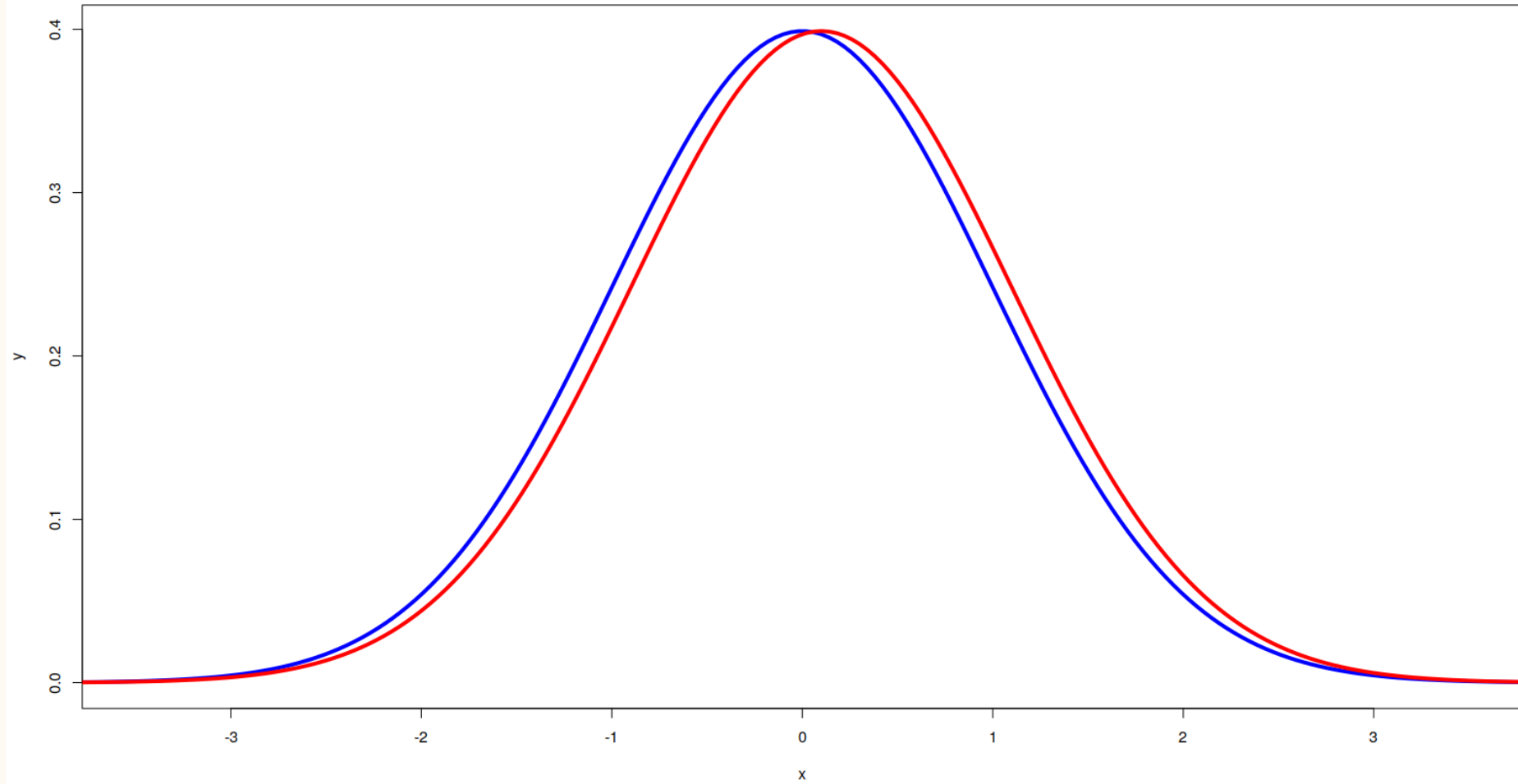
```
95 percent confidence interval:
```

```
-2.063833  4.618658
```

```
sample estimates:
```

```
mean of x mean of y
```

```
35.42667  34.14925
```



Questions

Take-home

- **(Blog)** [Too Much Knowledge? Security Beliefs and Protective Behaviors Among United States Internet Users](#) by Rick Wash and Emilee Rader
- **(Blog)** [Finland's Most-Wanted Hacker Nabbed in France](#)