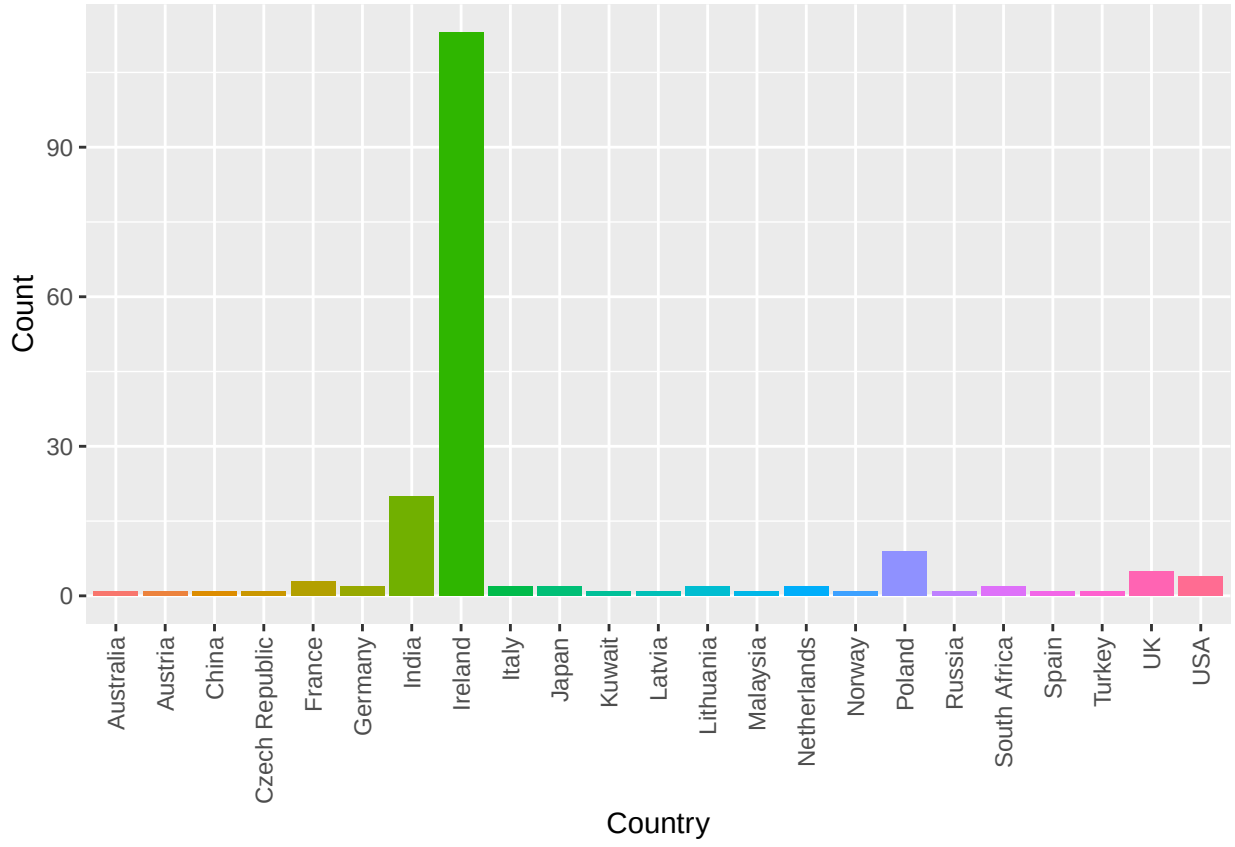
ST2001 Class Survey - Some Results

Carl Scarrott

9/11/2022

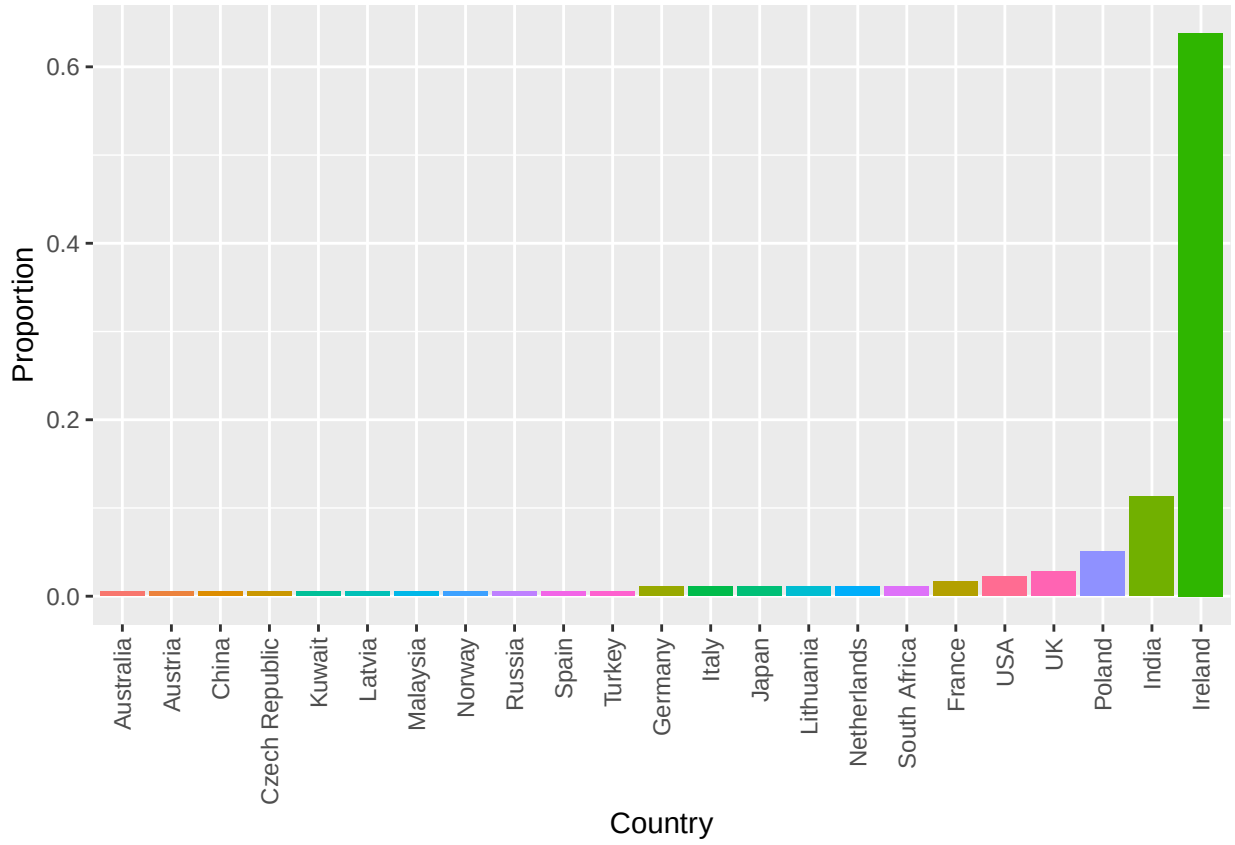
Q1. In which country were your born?

```
library(tidyverse)

survey.data = read.csv("ST2001wrangled.csv")

survey.data %>% filter(!is.na(country)) %>%
  mutate(country = factor(country)) %>%
  ggplot(aes(x = country, y = (..count..), fill = country)) +
  geom_bar() + labs(y = "Count", x = "Country") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1),
        legend.position = "none")
```



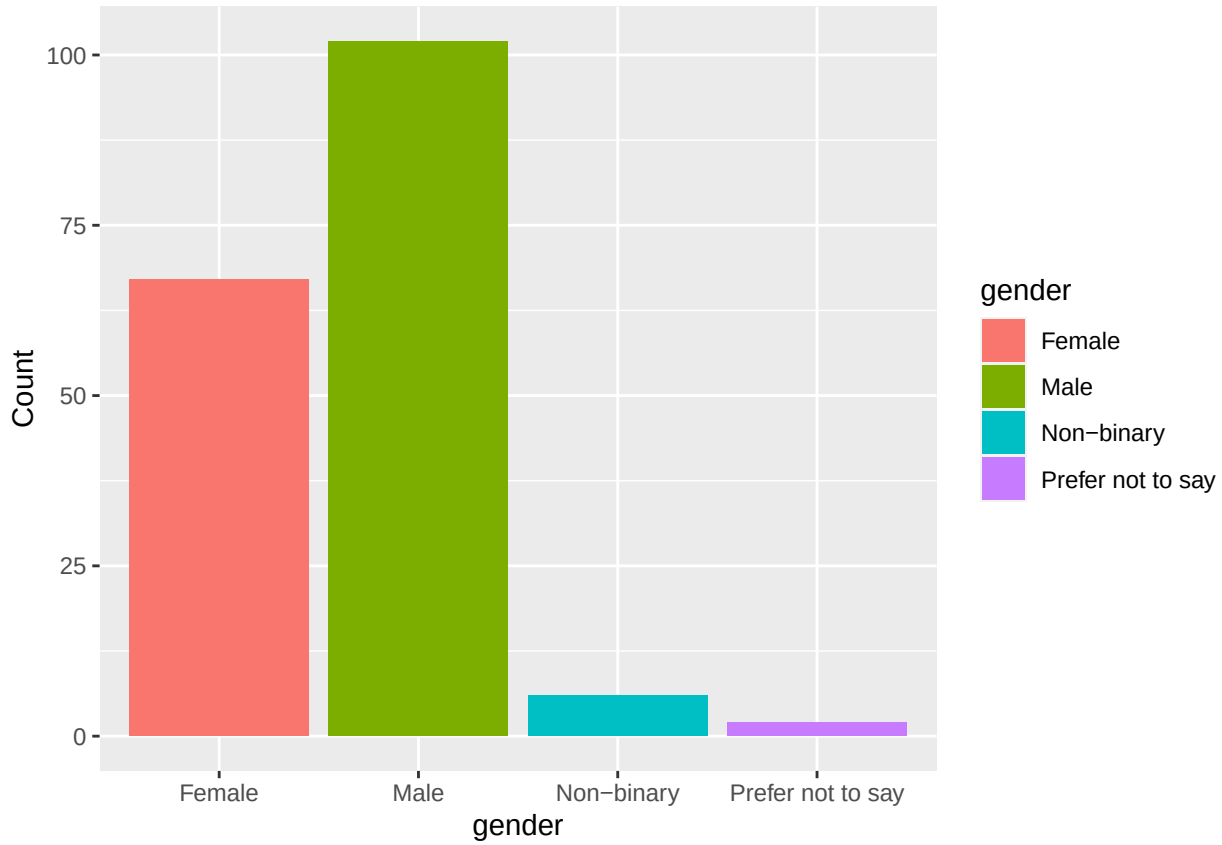

```
survey.data %>% filter(!is.na(country)) %>%
  mutate(country = factor(country)) %>%
```

```
  ggplot(aes(x = reorder(country, country, length), y = (..count..) / sum(..count..), fill = country)) +
  geom_bar() + labs(y = "Proportion", x = "Country") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1),
        legend.position = "none")
```

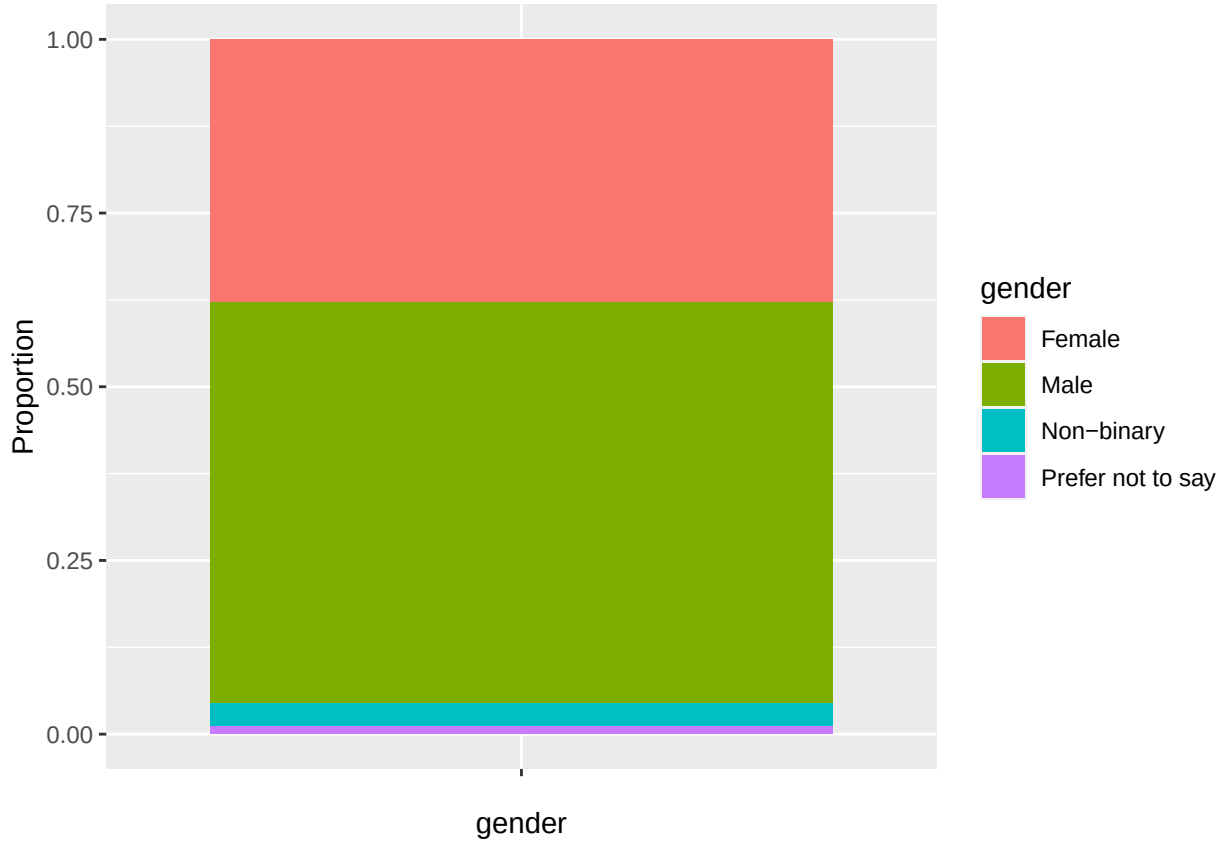


Q2. Gender?

```
survey.data %>% mutate(gender = factor(gender)) %>%
  ggplot(aes(x = gender, y = (..count..), fill = gender)) +
  geom_bar() + labs(y = "Count", x = "gender")
```

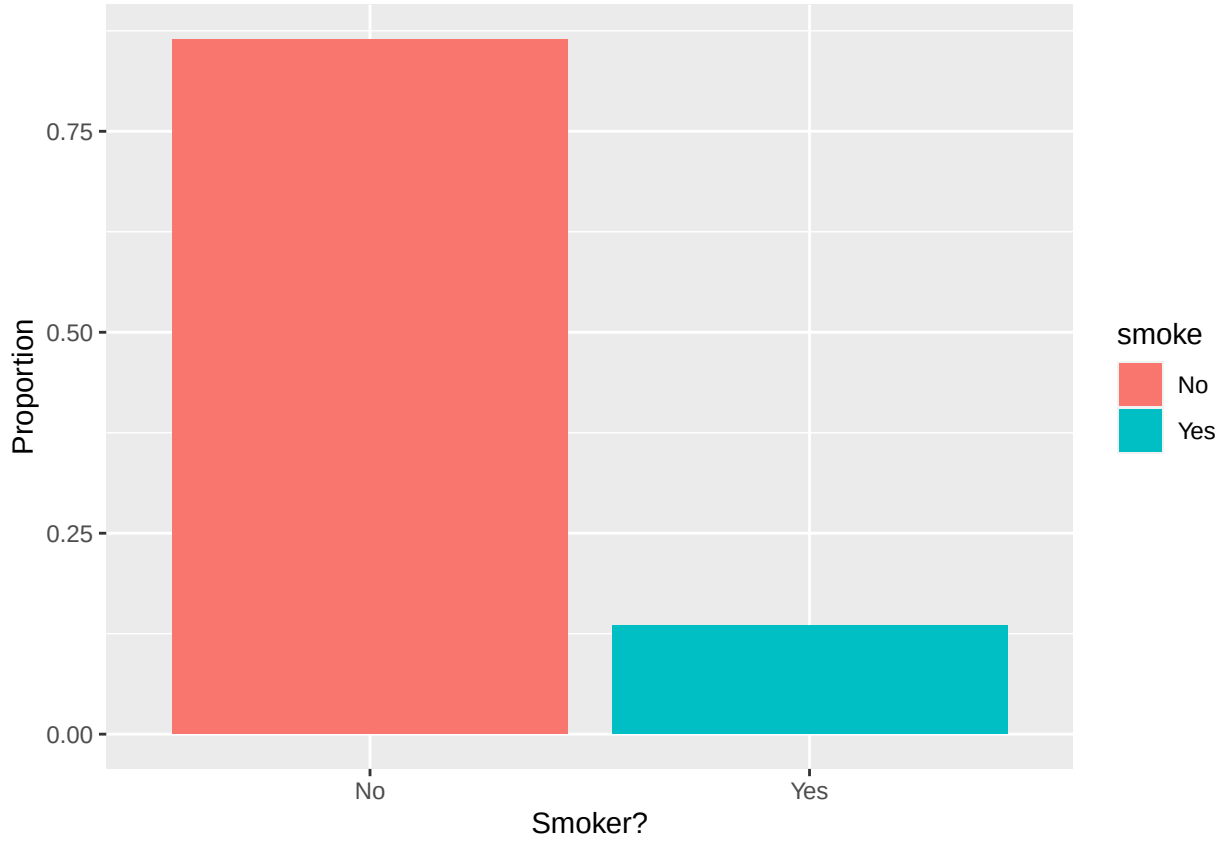



```
survey.data %>% mutate(gender = factor(gender)) %>%
  ggplot(aes(x = "", y = (..count..) / sum(..count..), fill = gender)) +
  geom_bar() + labs(y = "Proportion", x = "gender")
```



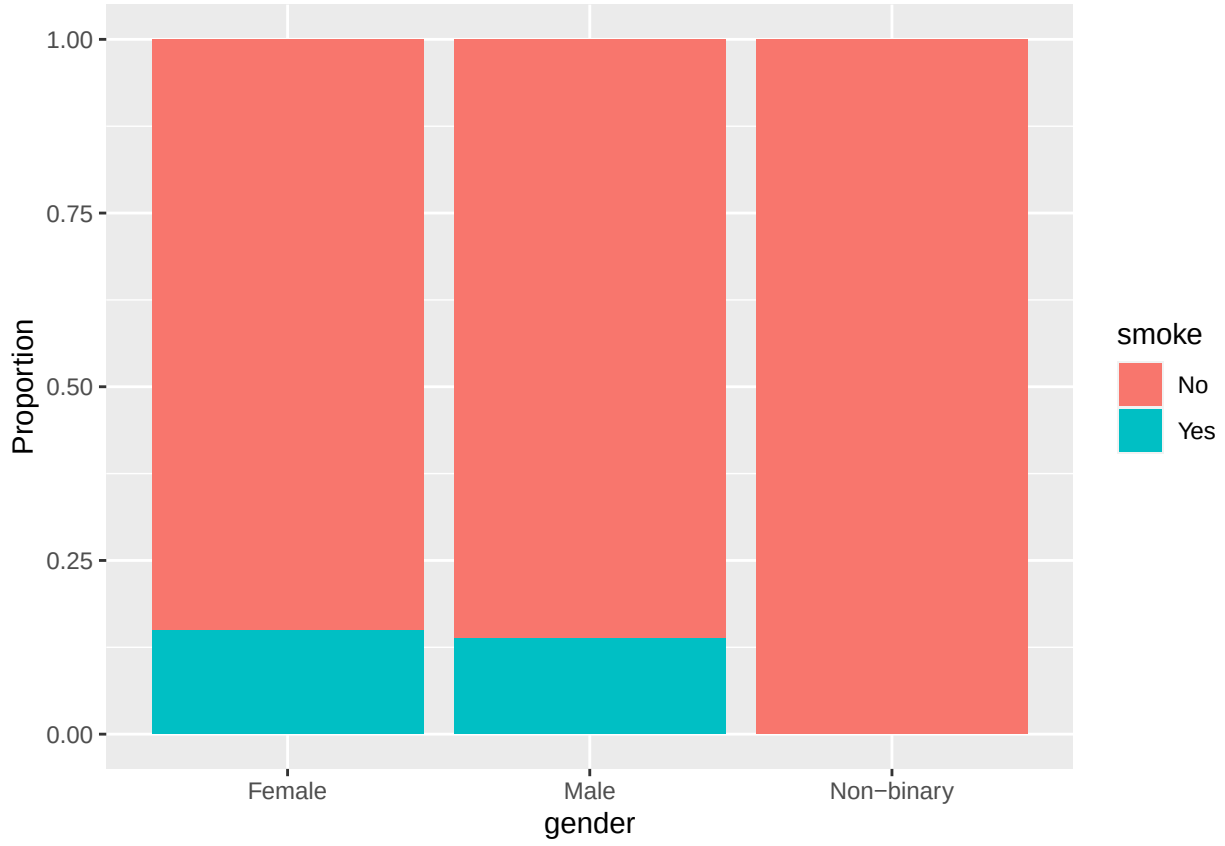
Q3. Do you smoke?

```
survey.data %>% mutate(smoke = factor(smoke)) %>%
  ggplot(aes(x = smoke, y = (..count..) / sum(..count..), fill = smoke)) +
  geom_bar() + labs(y = "Proportion", x = "Smoker?")
```



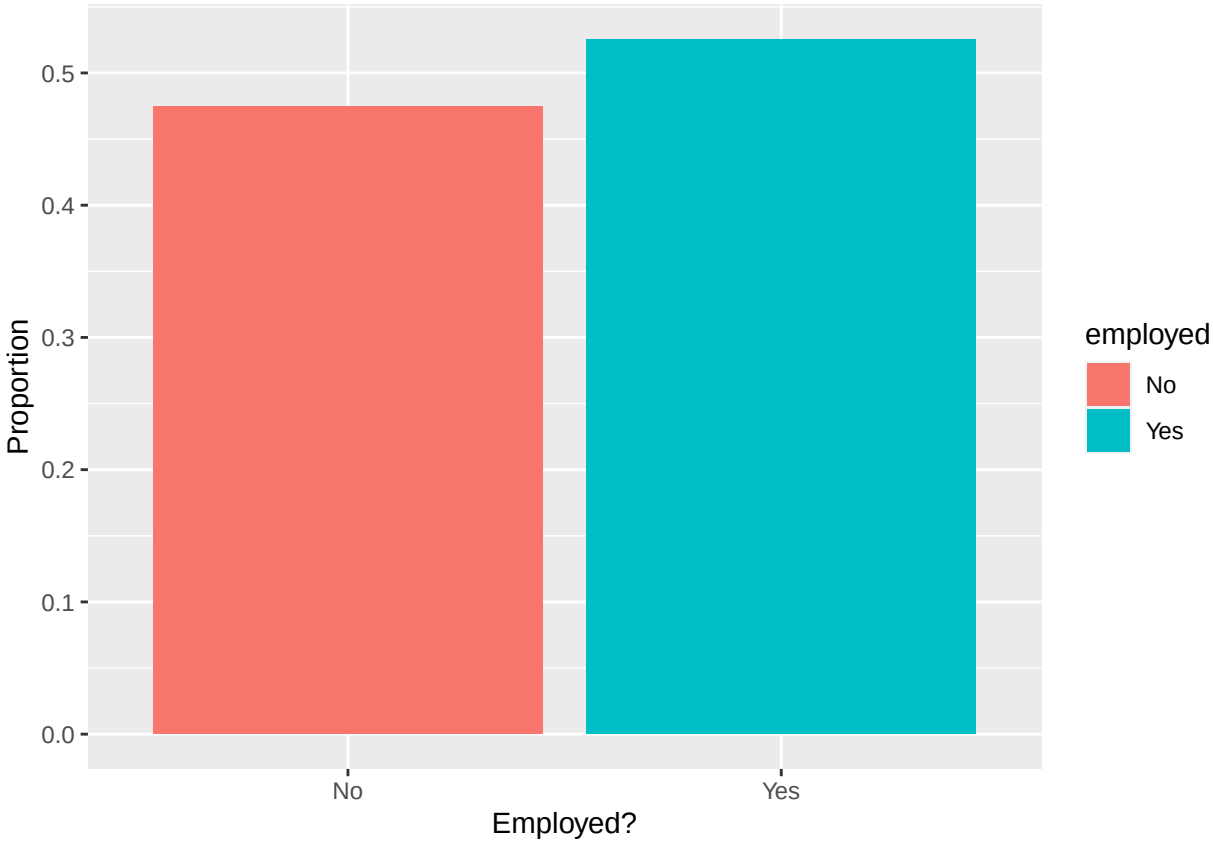
Smoking aggregated by gender.

```
survey.data %>% mutate(smoke = factor(smoke), gender = factor(gender)) %>%
  filter(gender != "Prefer not to say") %>%
  ggplot(aes(x = gender, y = (..count..) / sum(..count..), fill = smoke)) +
  geom_bar(position = "fill") + labs(y = "Proportion", x = "gender")
```



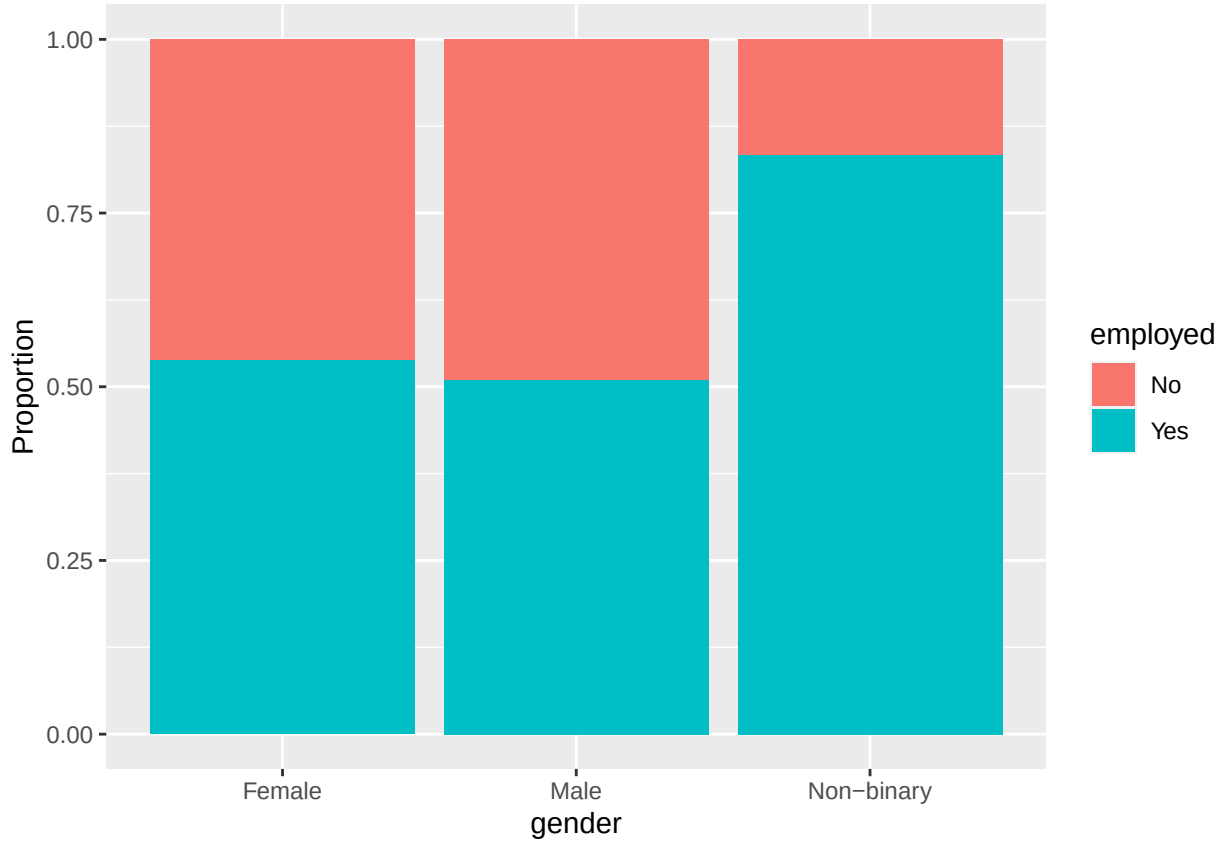
Q4. Are you employed?

```
survey.data %>% mutate(employed = factor(employed)) %>%
  ggplot(aes(x = employed, y = (..count..) / sum(..count..), fill = employed)) +
  geom_bar() + labs(y = "Proportion", x = "Employed?")
```



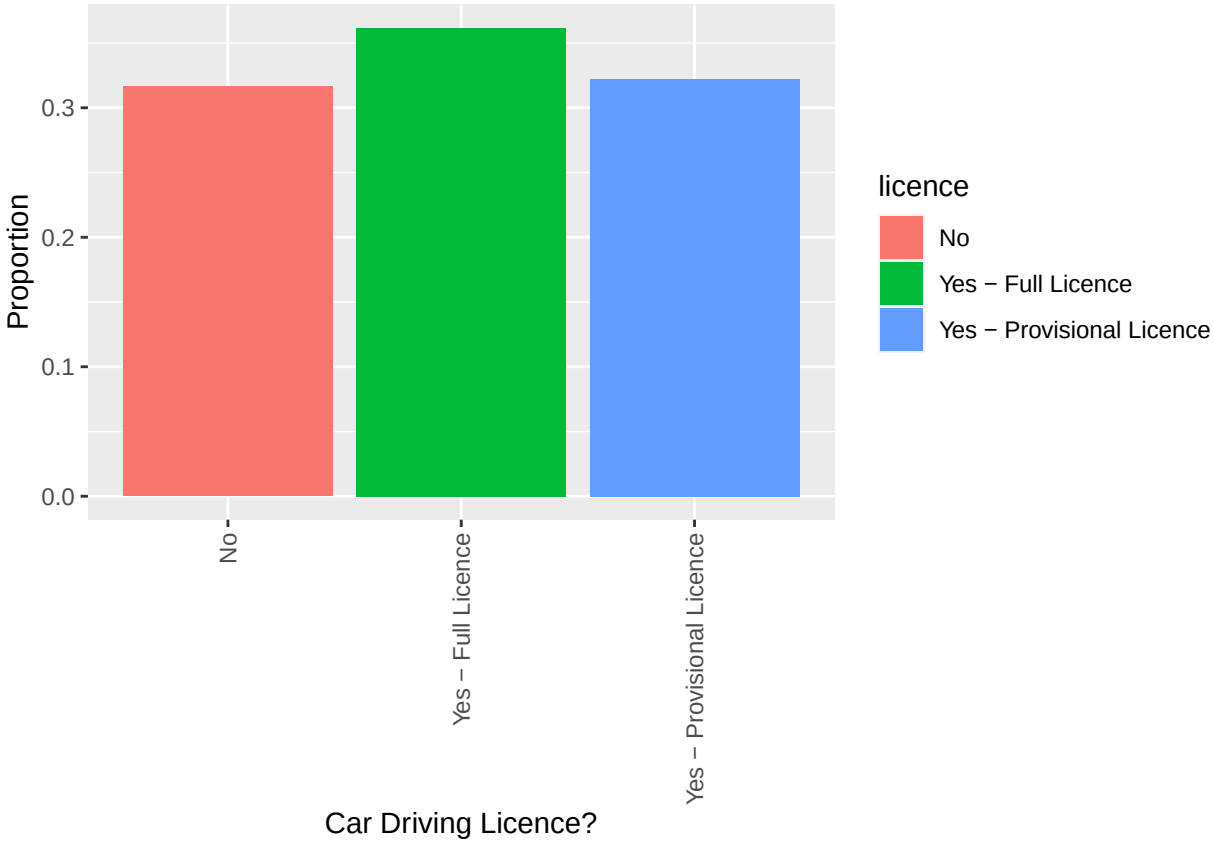
Employment status by gender.

```
survey.data %>% mutate(employed = factor(employed), gender = factor(gender)) %>%
  filter(gender != "Prefer not to say") %>%
  ggplot(aes(x = gender, y = (..count..) / sum(..count..), fill = employed)) +
  geom_bar(position = "fill") + labs(y = "Proportion", x = "gender")
```



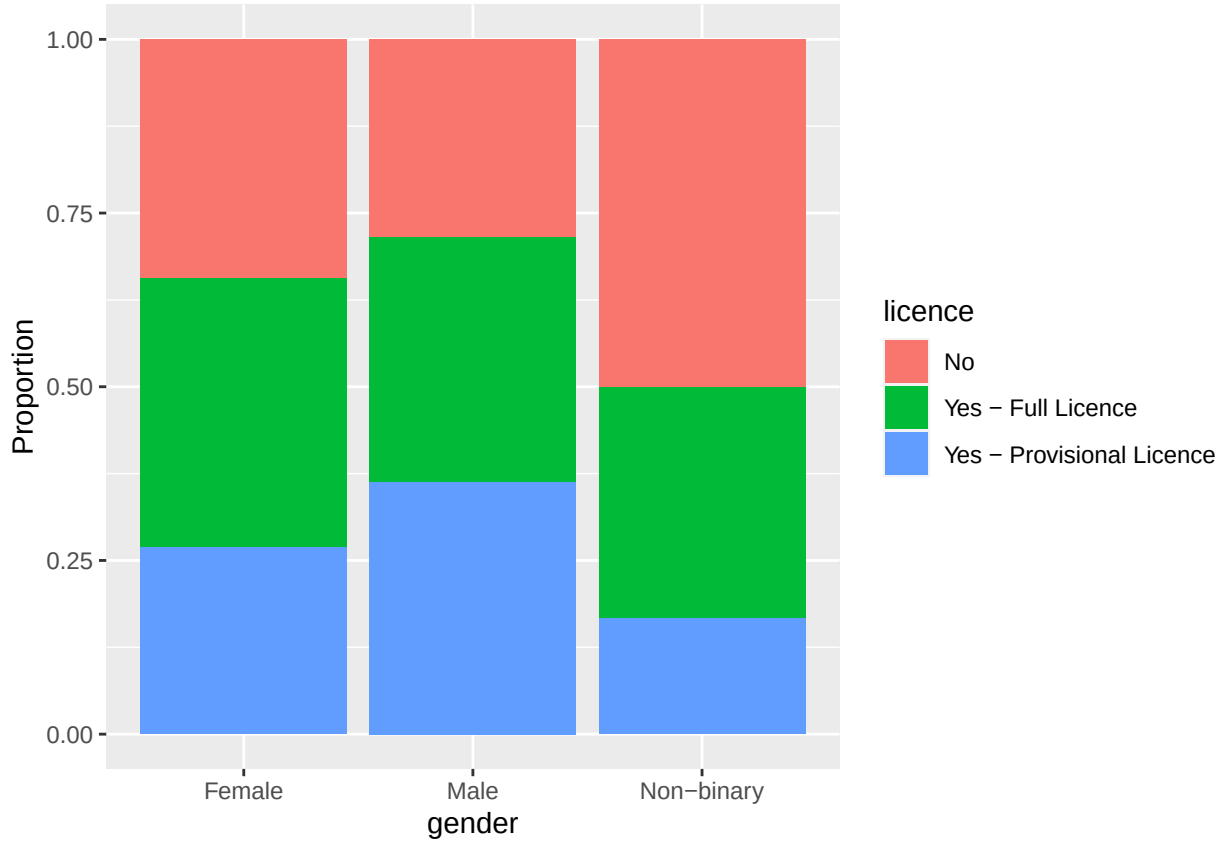
Q5. Do you have a car driving licence?

```
survey.data %>% mutate(licence = factor(licence)) %>%
  ggplot(aes(x = licence, y = (..count..) / sum(..count..), fill = licence)) +
  geom_bar() + labs(y = "Proportion", x = "Car Driving Licence?") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
```



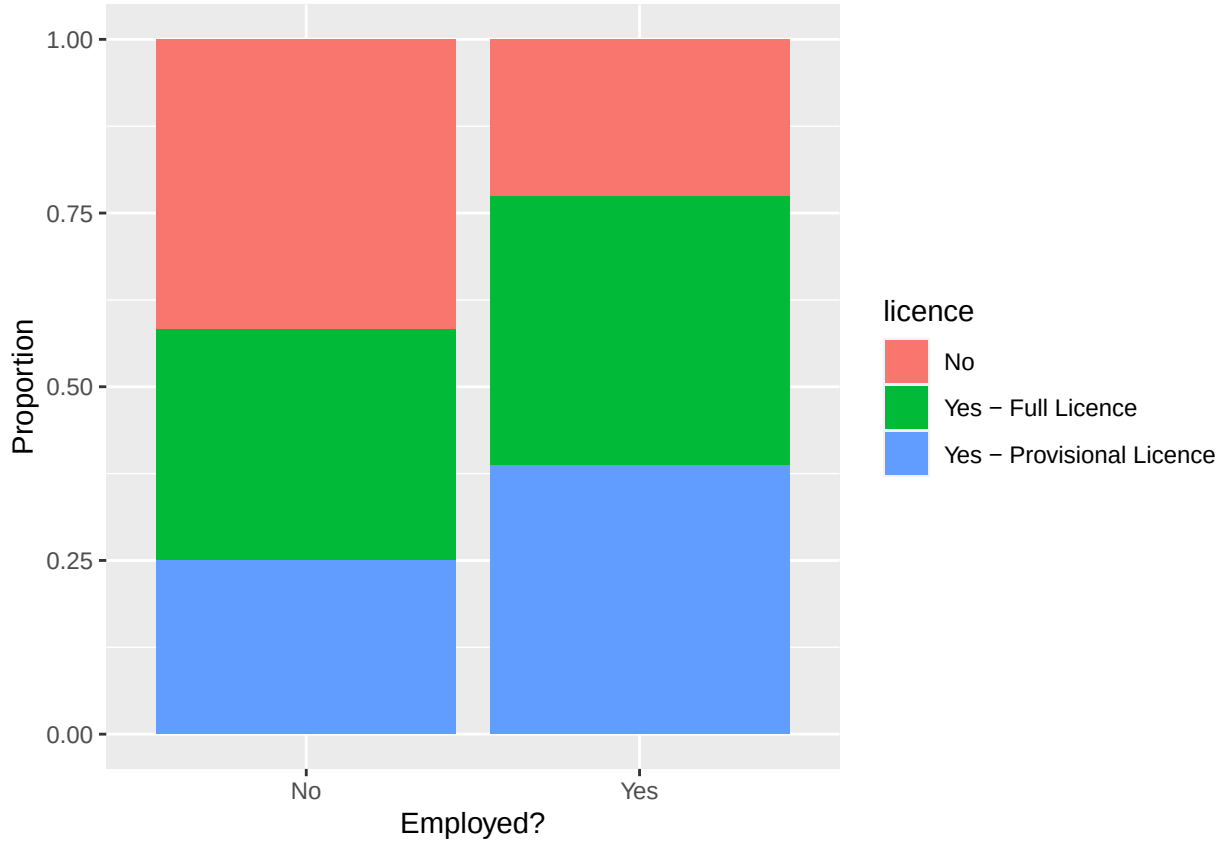
Car driving licence by gender.

```
survey.data %>% mutate(licence = factor(licence), gender = factor(gender)) %>%
  filter(gender != "Prefer not to say") %>%
  ggplot(aes(x = gender, y = (..count..) / sum(..count..), fill = licence)) +
  geom_bar(position = "fill") + labs(y = "Proportion", x = "gender")
```



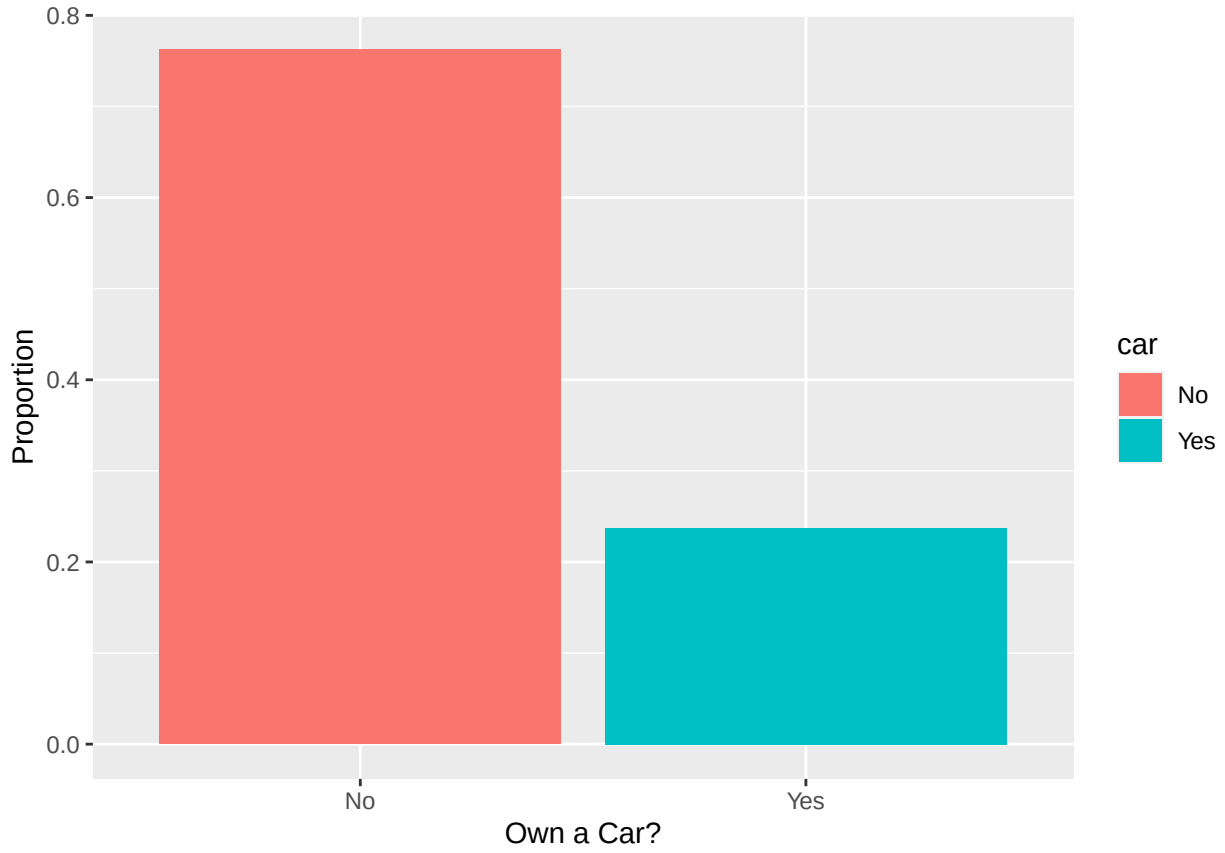
Car driving licence by employment status.

```
survey.data %>% mutate(licence = factor(licence), employed = factor(employed)) %>%
  ggplot(aes(x = employed, y = (..count..) / sum(..count..), fill = licence)) +
  geom_bar(position = "fill") + labs(y = "Proportion", x = "Employed?")
```



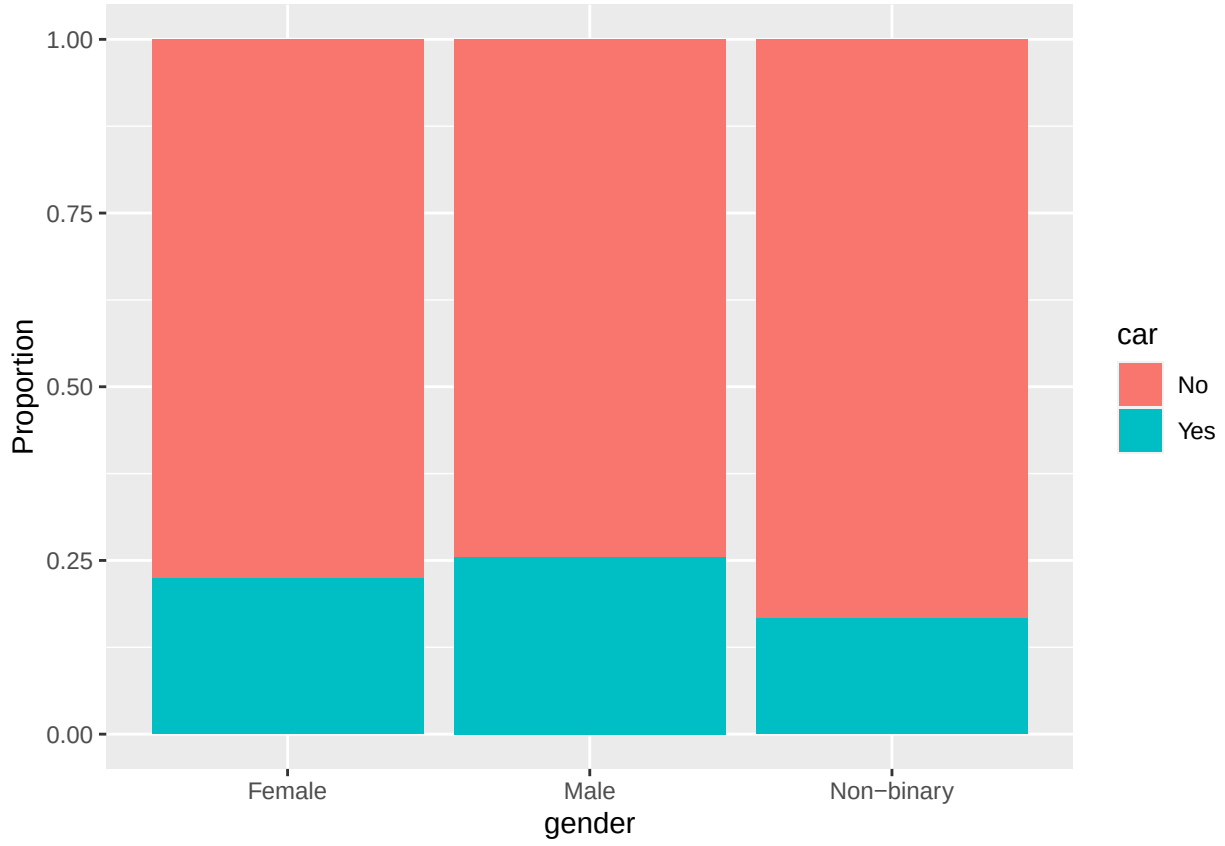
Q6. Do you own a car?

```
survey.data %>% mutate(car = factor(car)) %>%
  ggplot(aes(x = car, y = (..count..) / sum(..count..), fill = car)) +
  geom_bar() + labs(y = "Proportion", x = "Own a Car?")
```



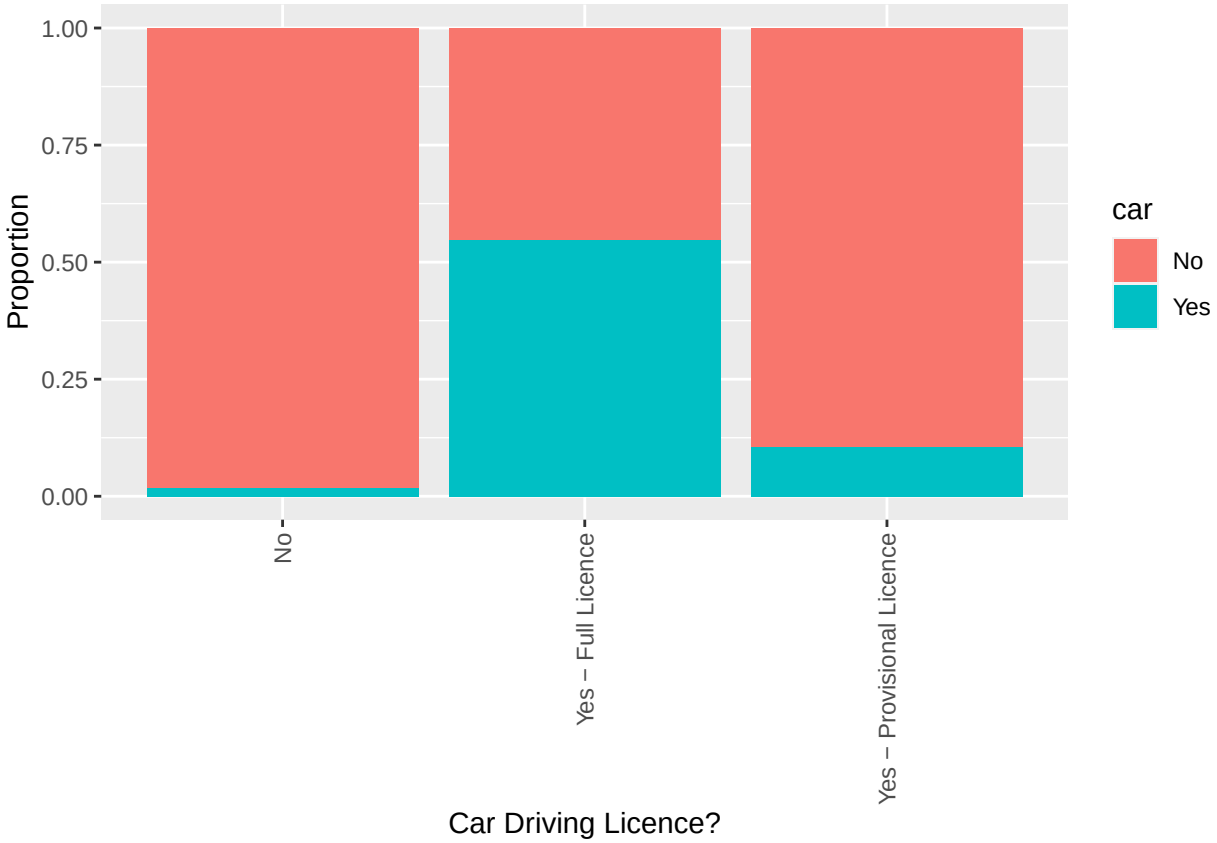
Car ownership by gender.

```
survey.data %>% mutate(car = factor(car), gender = factor(gender)) %>%
  filter(gender != "Prefer not to say") %>%
  ggplot(aes(x = gender, y = (..count..) / sum(..count..), fill = car)) +
  geom_bar(position = "fill") + labs(y = "Proportion", x = "gender")
```



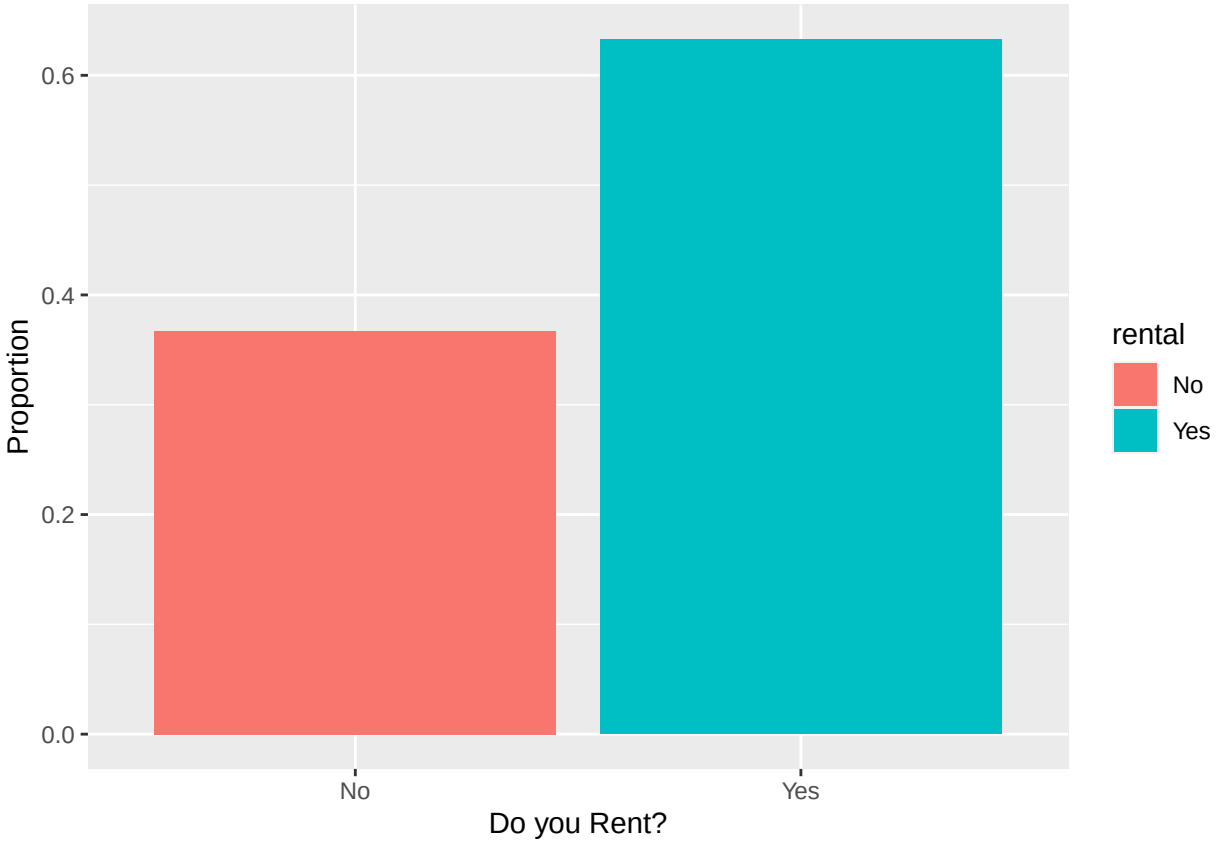
Car ownership by driving licence status.

```
survey.data %>% mutate(car = factor(car), licence = factor(licence)) %>%
  ggplot(aes(x = licence, y = (..count..) / sum(..count..), fill = car)) +
  geom_bar(position = "fill") + labs(y = "Proportion", x = "Car Driving Licence?") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
```



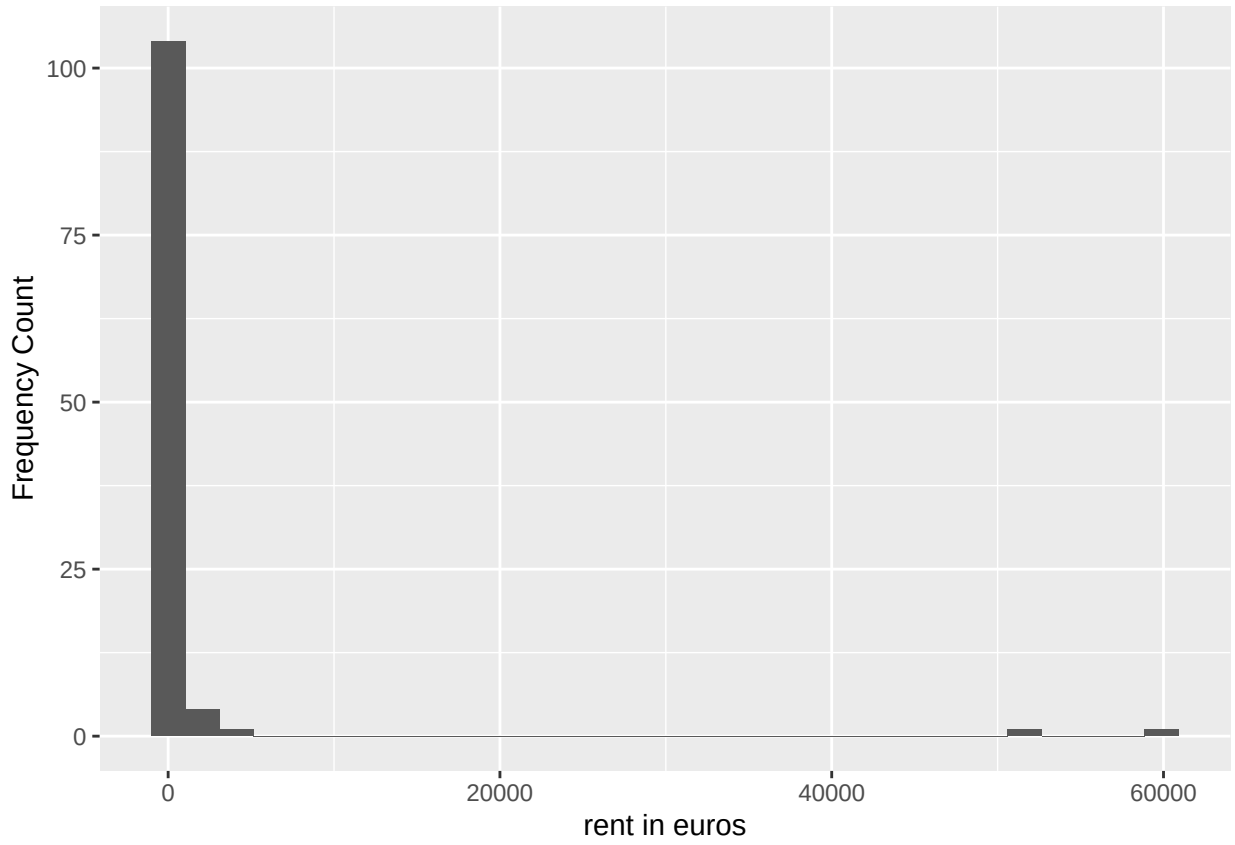
Q7. Do you rent accommodation?

```
survey.data %>% mutate(rental = factor(rental)) %>%
  ggplot(aes(x = rental, y = (..count..) / sum(..count..), fill = rental)) +
  geom_bar() + labs(y = "Proportion", x = "Do you Rent?")
```



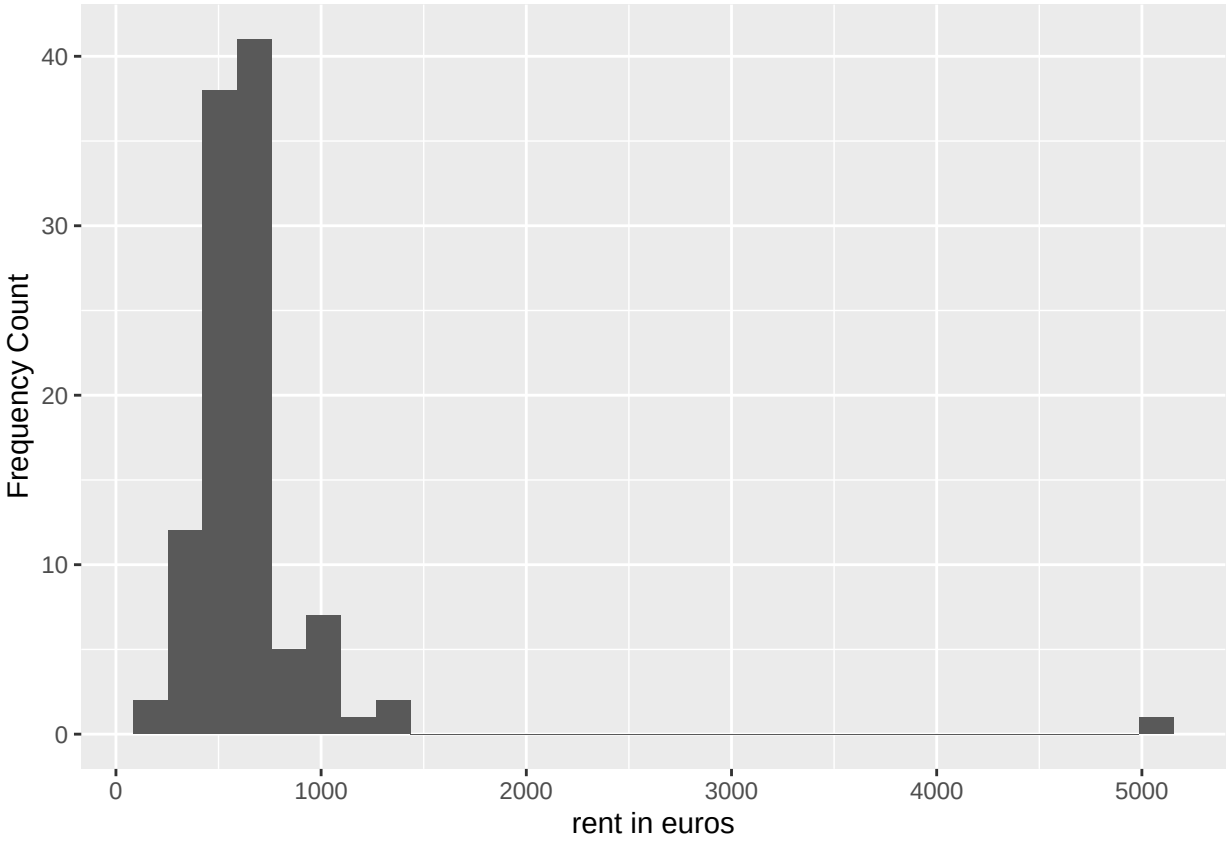
Q8. If so, what is you monthly rent (in €)?

```
survey.data %>% ggplot(aes(x = rent)) +
  geom_histogram() + labs(y = "Frequency Count", x = "rent in euros")
```

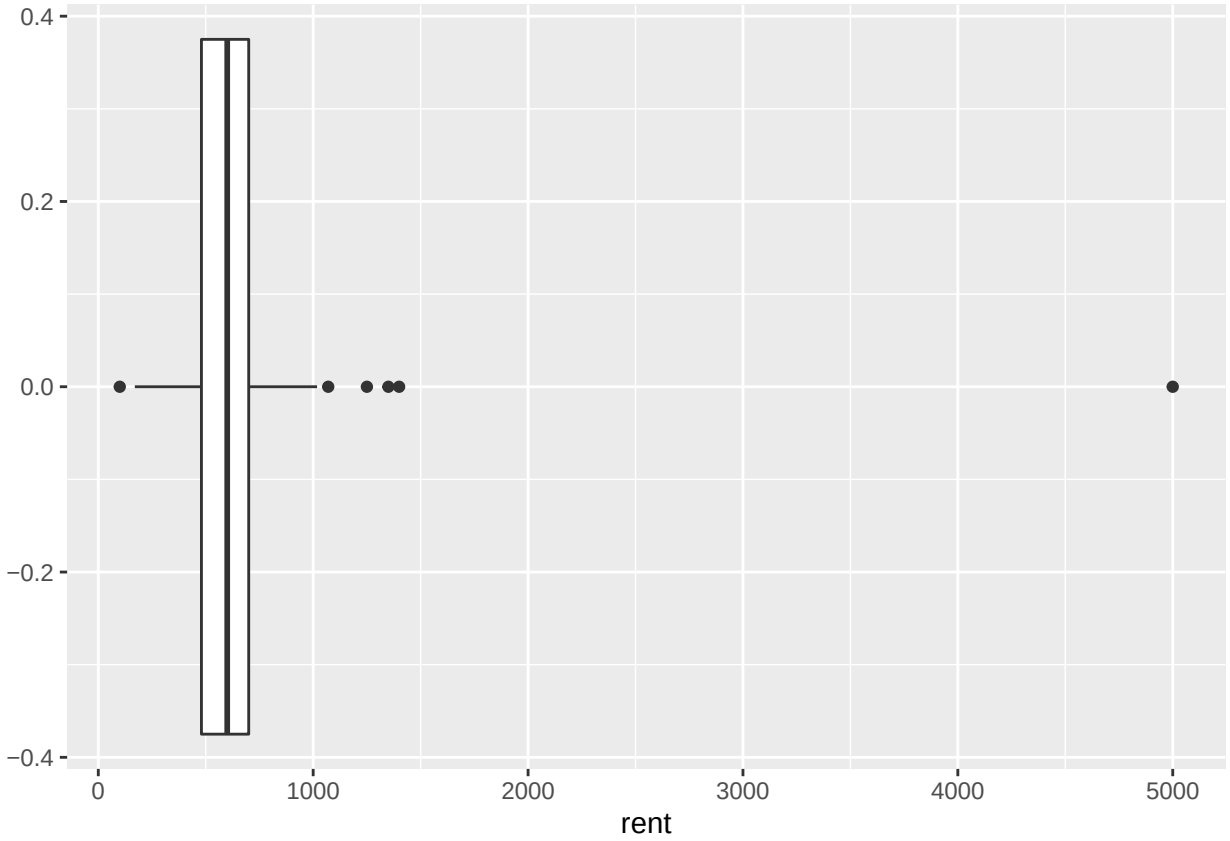


Let's remove the outlying rent values. Either they are very rich or possibly these are temporary accommodation high rates and so they may not be representative, or they were typoes!

```
survey.data %>% filter(rent < 10000) %>%
  ggplot(aes(x = rent)) +
  geom_histogram() + labs(y = "Frequency Count", x = "rent in euros")
```

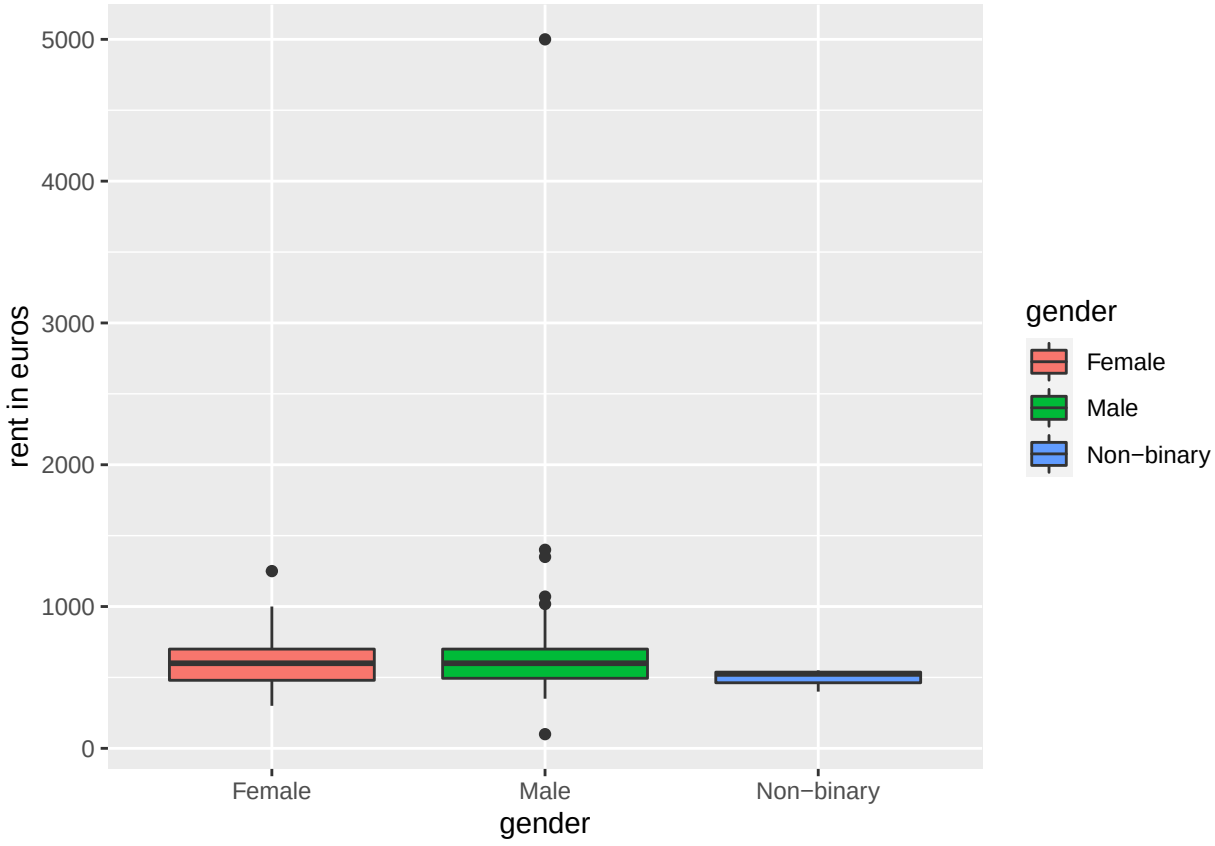



```
survey.data %>% filter(rent < 10000) %>%
  ggplot(aes(x = rent)) +
  geom_boxplot()
```



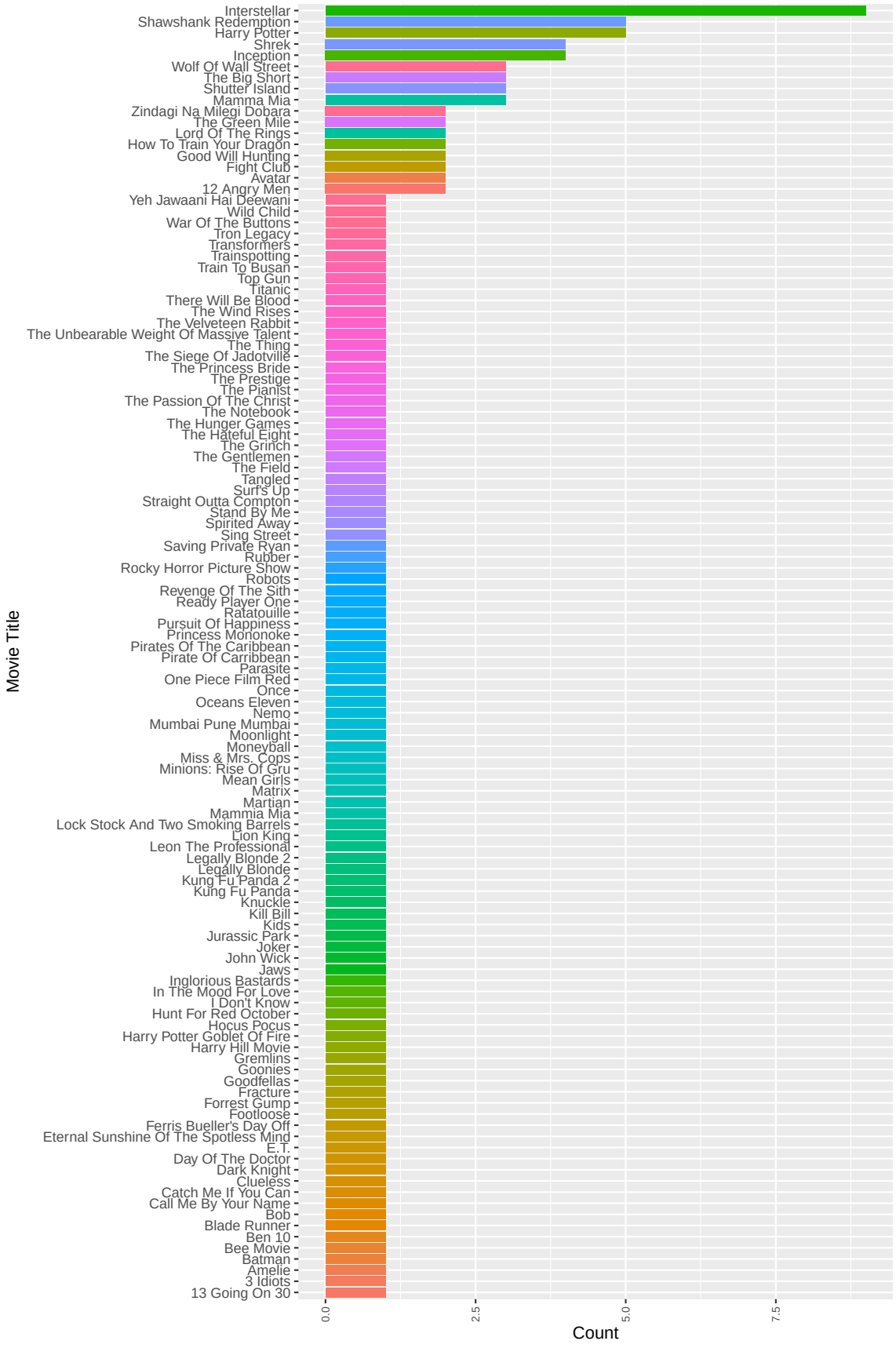
Difference in rent for each gender is

```
survey.data %>% filter(rent < 10000, gender != "Prefer not to say") %>%
  ggplot(aes(x = gender, y = rent, fill = gender)) +
  geom_boxplot() + labs(y = "rent in euros")
```



Q9. What is your favourite movie?

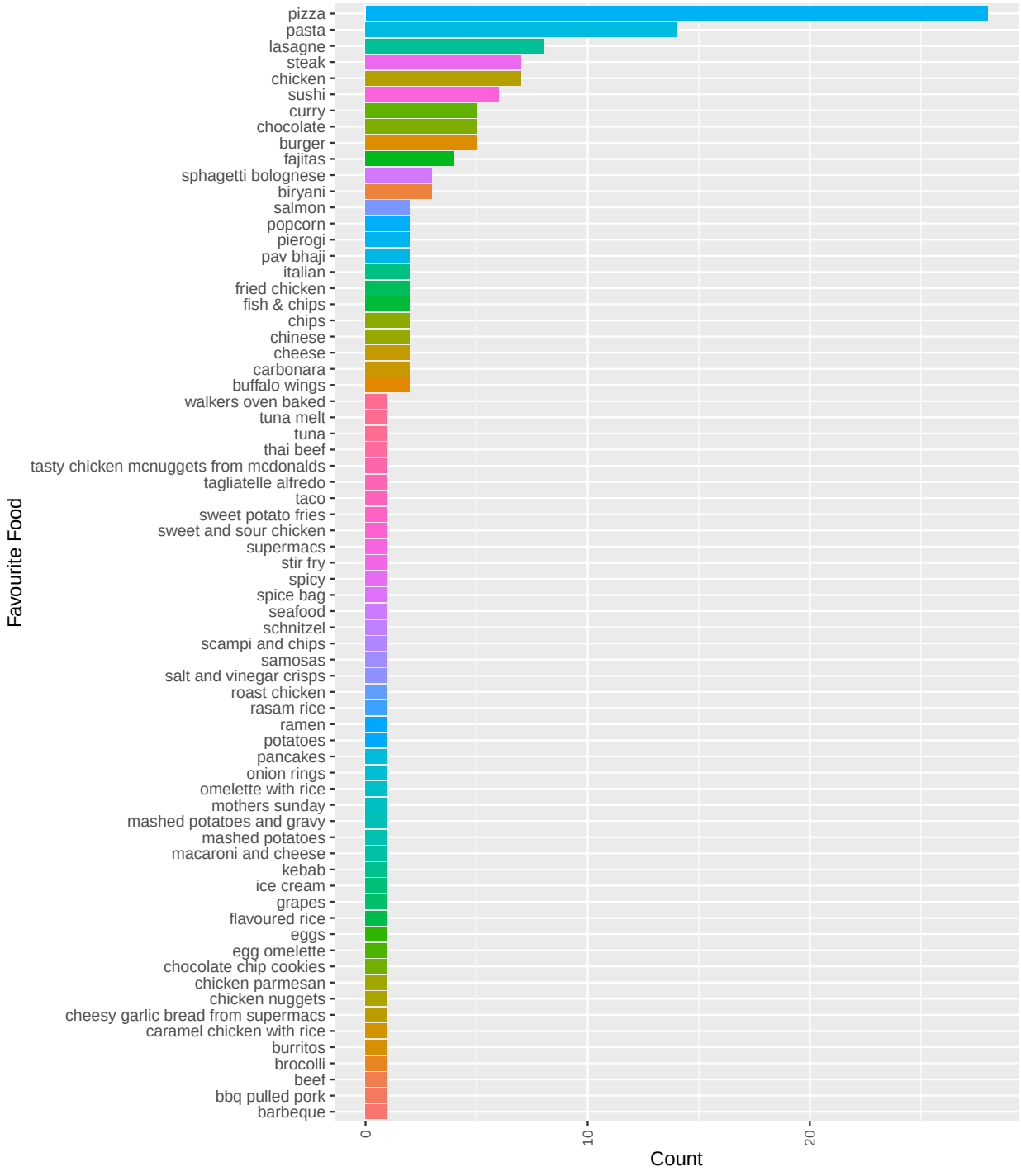
```
survey.data %>% mutate(movie = factor(movie)) %>% select(movie) %>% na.omit(movie) %>%
  ggplot(aes(x = reorder(movie, movie, length), y = (..count..), fill = movie)) +
  geom_bar() + labs(y = "Count", x = "Movie Title") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1, size = 7)) +
  theme(legend.position = "none") + coord_flip()
```


Interstellar
Shawshank Redemption
Harry Potter
Shrek
Inception
Wolf Of Wall Street
The Big Short
Shutter Island
Mamma Mia
Zindagi Na Milegi Dobara
The Green Mile
Lord Of The Rings
How To Train Your Dragon
Good Will Hunting
Fight Club
Avatar
12 Angry Men
Yeh Jawaani Hai Deewani
Wild Child
War Of The Buttons
Tron Legacy
Transformers
Trainspotting
Train To Busan
Top Gun
Titanic
There Will Be Blood
The Wind Rises
The Velveteen Rabbit
The Unbearable Weight Of Massive Talent
The Thing
The Siege Of Jadotville
The Princess Bride
The Prestige
The Pianist
The Passion Of The Christ
The Notebook
The Hunger Games
The Hateful Eight
The Grinch
The Gentlemen
The Field
Tangled
Surf's Up
Straight Outta Compton
Stand By Me
Spirited Away
Sing Street
Saving Private Ryan
Rubber
Rocky Horror Picture Show
Robots
Revenge Of The Sith
Ready Player One
Ratatouille
Pursuit Of Happiness
Princess Mononoke
Pirates Of The Caribbean
Pirate Of Carribbean
Parasite
One Piece Film Red
Once
Oceans Eleven
Nemo
Mumbai Pune Mumbai
Moonlight
Moneyball
Miss & Mrs. Cops
Minions: Rise Of Gru
Mean Girls
Matrix
Martian
Mammia Mia
Lock Stock And Two Smoking Barrels
Lion King
Leon The Professional
Legally Blonde 2
Legally Blonde
Kung Fu Panda 2
Kung Fu Panda
Knuckle
Kill Bill
Kids
Jurassic Park
Joker
John Wick
Jaws
Inglorious Bastards
In The Mood For Love
I Don't Know
Hunt For Red October
Hocus Pocus
Harry Potter Goblet Of Fire
Harry Hill Movie
Gremlins
Goonies
Goodfellas
Fracture
Forrest Gump
Footloose
Ferris Bueller's Day Off
Eternal Sunshine Of The Spotless Mind
E.T.
Day Of The Doctor
Dark Knight
Clueless
Catch Me If You Can
Call Me By Your Name
Bob
Blade Runner
Ben 10
Bee Movie
Batman
Amelie
3 Idiots
13 Going On 30
Movie Title
0.0
2.5
5.0
7.5
Count

Q10. What is your favourite food?

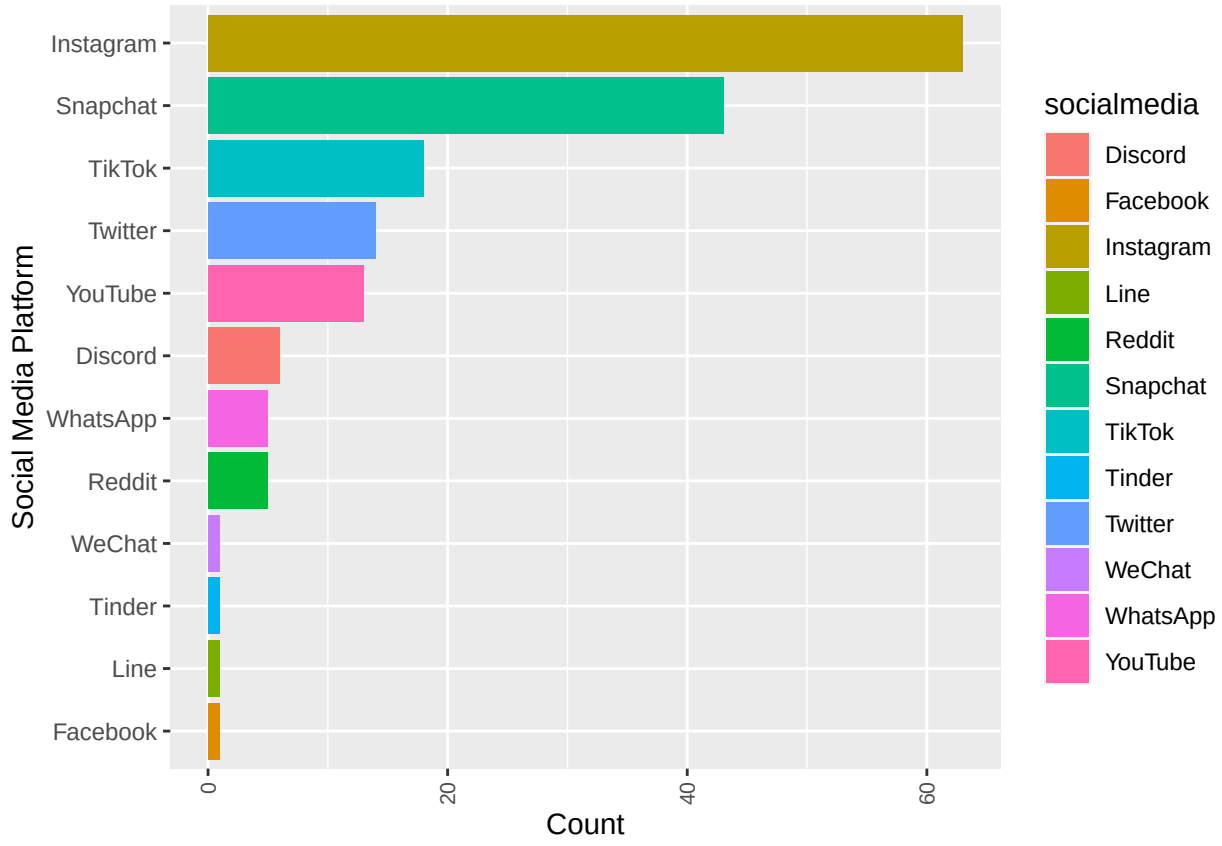
Note that some responses have been simplified for display and aggregation purposes.

```
survey.data %>% mutate(food = factor(food)) %>% select(food) %>% na.omit(food) %>%
  ggplot(aes(x = reorder(food, food, length), y = (..count..), fill = food)) +
  geom_bar() + labs(y = "Count", x = "Favourite Food") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1, size = 8)) +
  theme(legend.position = "none") + coord_flip()
```



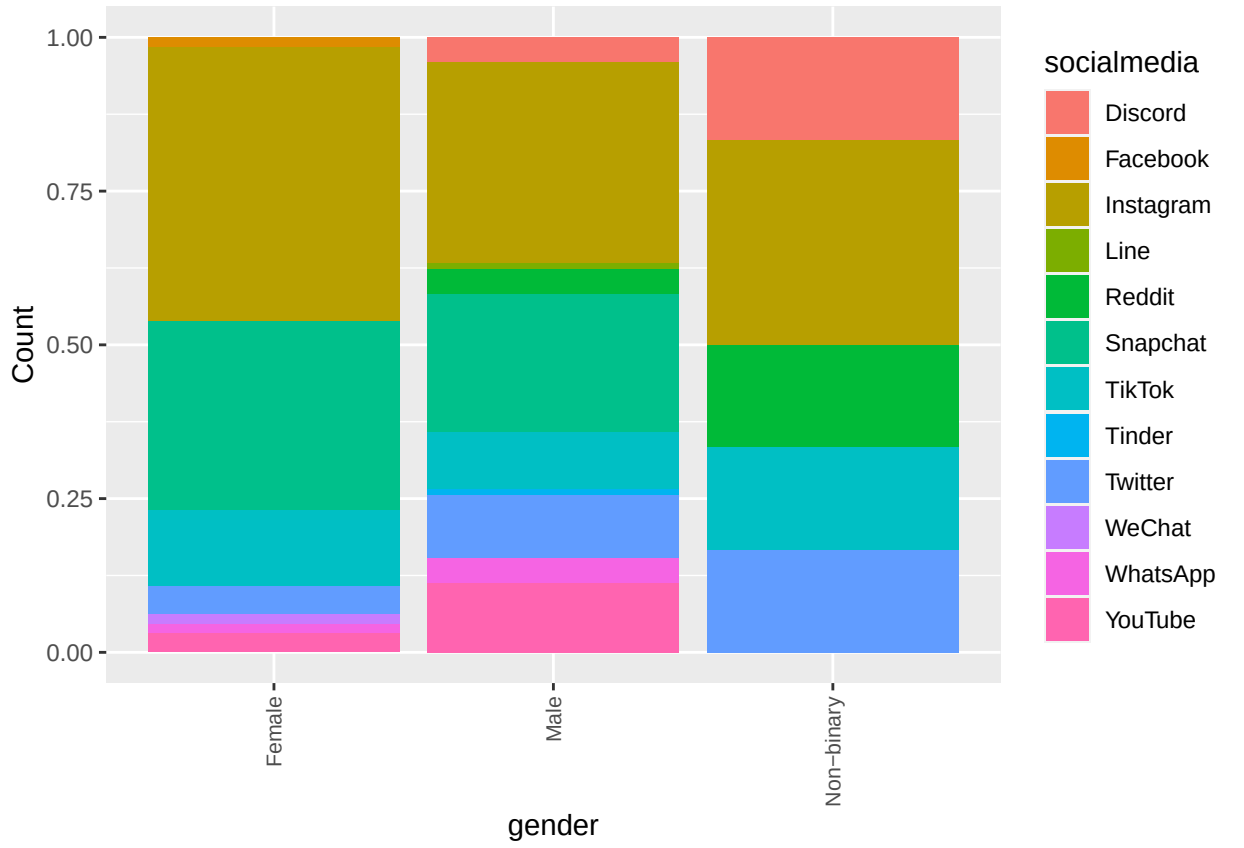
Q11. What social media platform do you the most?

```
survey.data %>% filter(!is.na(socialmedia)) %>%
  mutate(socialmedia = factor(socialmedia)) %>%
  ggplot(aes(x = reorder(socialmedia, socialmedia, length), y = (..count..), fill = socialmedia)) +
  geom_bar() + labs(y = "Count", x = "Social Media Platform") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1, size = 8)) + coord_flip()
```



Social media most used by gender.

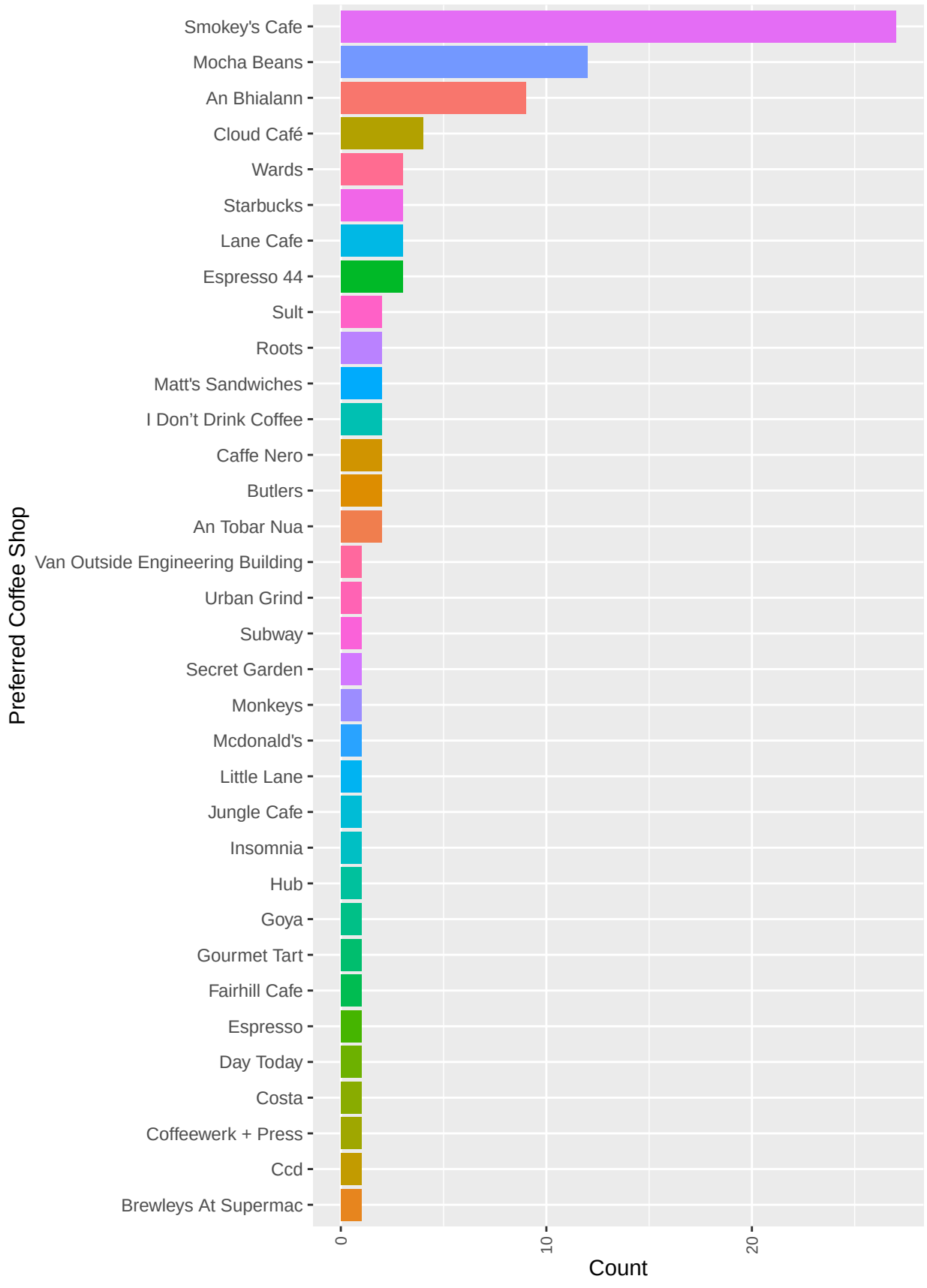
```
survey.data %>% filter(!is.na(socialmedia)) %>%
  mutate(socialmedia = factor(socialmedia), gender = factor(gender)) %>%
  filter(gender != "Prefer not to say") %>%
  ggplot(aes(x = gender, y = (..count..), fill = socialmedia)) +
  geom_bar(position = "fill") + labs(y = "Count", x = "gender") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1, size = 8))
```



Q12. Where is the best coffee on (or near) campus

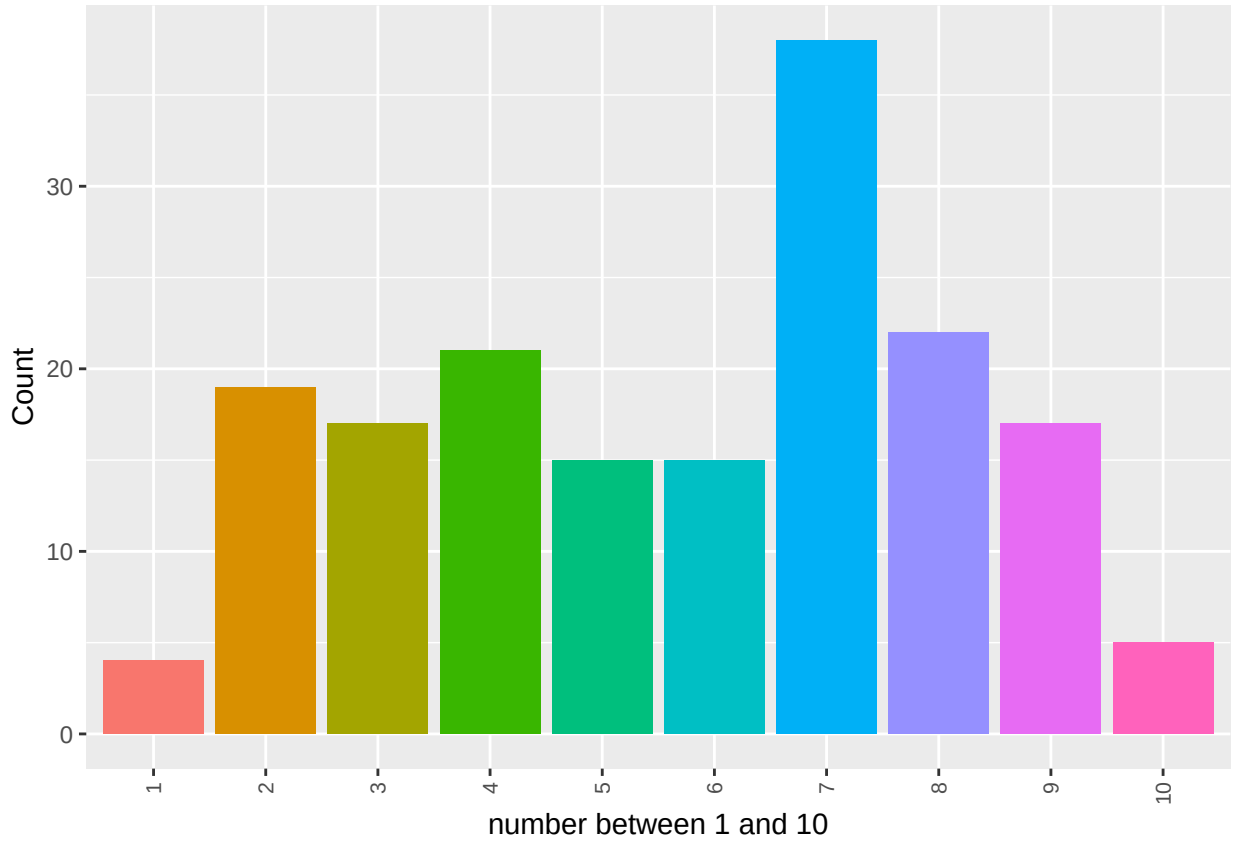
Response like "I don't drink coffee" and "I don't know" have been set to missing, so we can focus on preferred coffee shops.

```
survey.data %>% mutate(coffee = factor(coffee))  %>% select(coffee) %>% na.omit(coffee) %>%
  ggplot(aes(x = reorder(coffee, coffee, length), y = (..count..), fill = coffee)) +
  geom_bar() + labs(y = "Count", x = "Preferred Coffee Shop") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1, size = 8),
        legend.position = "none") + coord_flip()
```


Preferred Coffee Shop
Smokey's Cafe
Mocha Beans
An Bhialann
Cloud Café
Wards
Starbucks
Lane Cafe
Espresso 44
Sult
Roots
Matt's Sandwiches
I Don't Drink Coffee
Caffe Nero
Butlers
An Tobar Nua
Van Outside Engineering Building
Urban Grind
Subway
Secret Garden
Monkeys
Mcdonald's
Little Lane
Jungle Cafe
Insomnia
Hub
Goya
Gourmet Tart
Fairhill Cafe
Espresso
Day Today
Costa
Coffeewerk + Press
Ccd
Brewleys At Supermac
0
10
20
Count

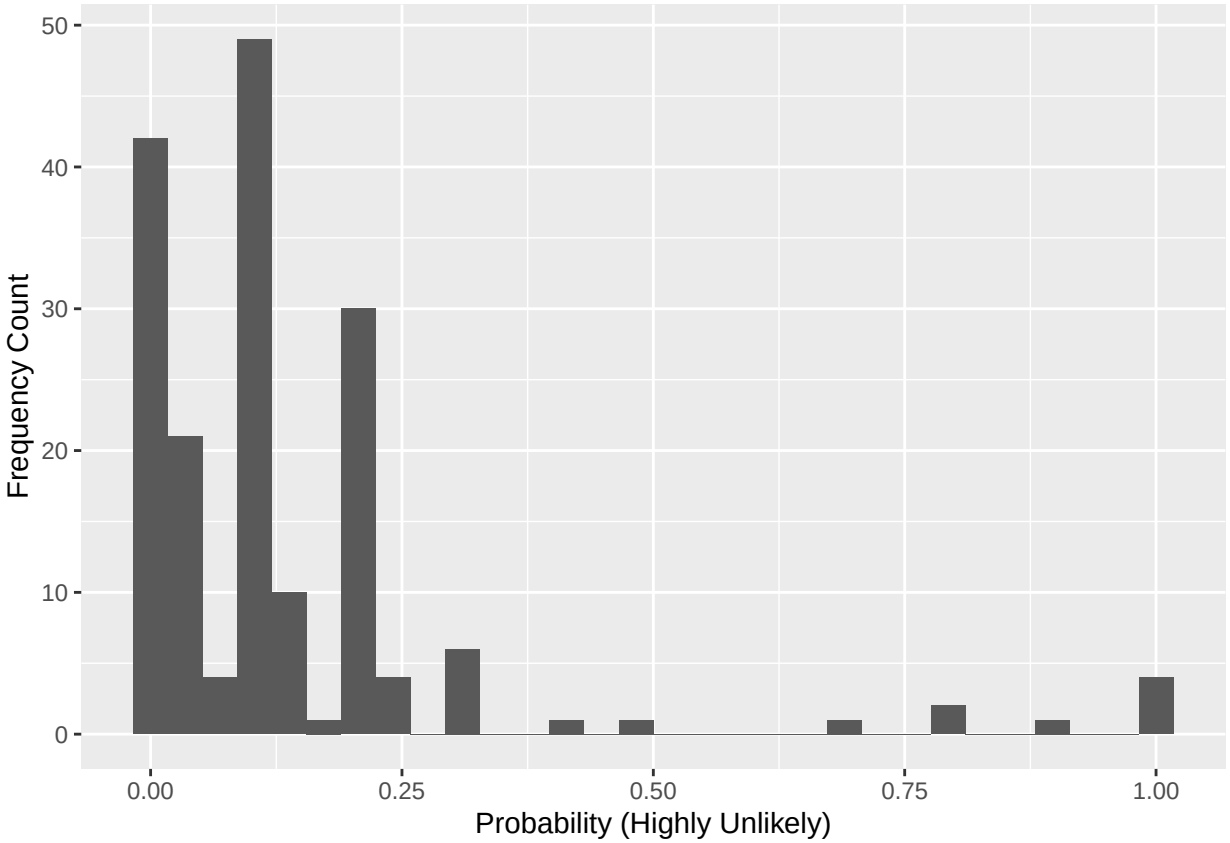
Q13. Pick a (whole) number at random between 1 and 10

```
survey.data %>% filter(!is.na(number)) %>% mutate(number = factor(number)) %>%
  ggplot(aes(x = number, y = (..count..), fill = number)) +
  geom_bar() + labs(y = "Count", x = "number between 1 and 10") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1, size = 8),
        legend.position = "none")
```



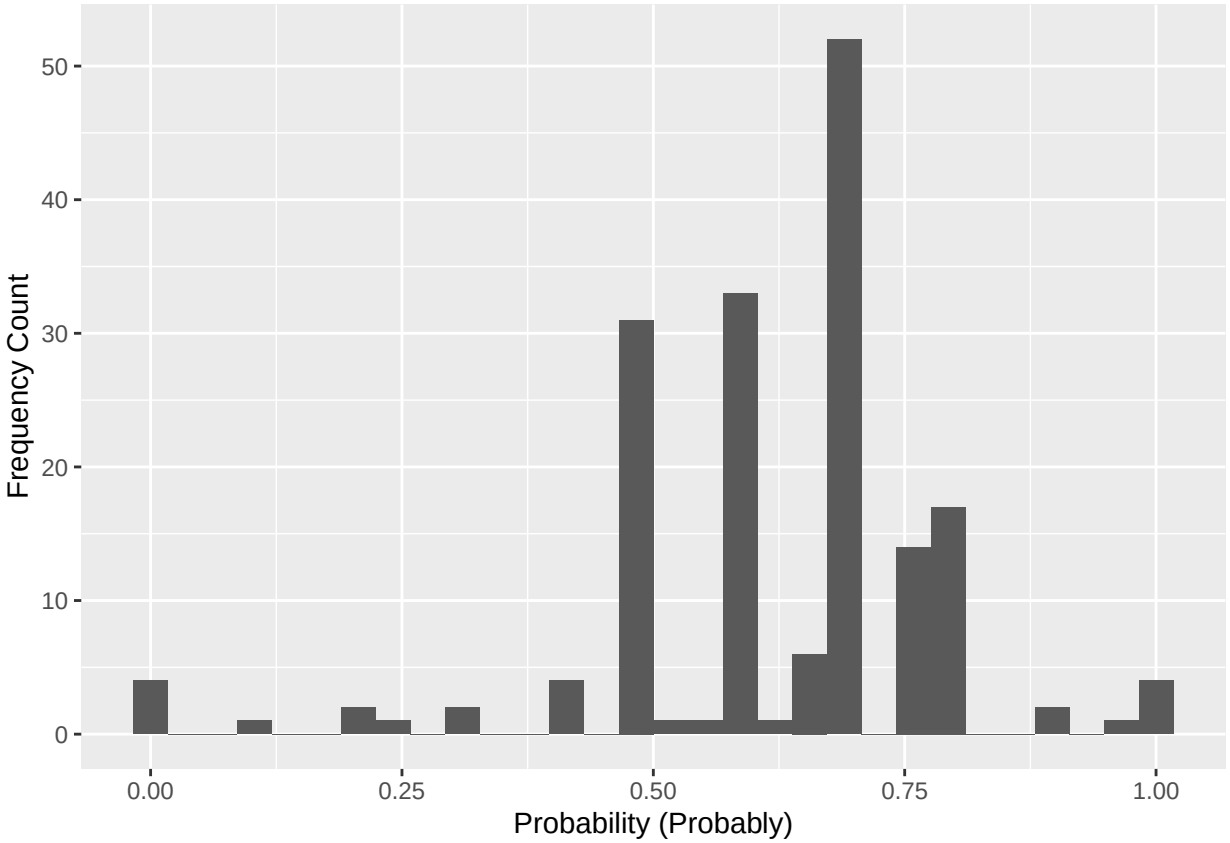
Q14. If you were told it was "Highly Unlikely" that a natural disaster was about to occur. What is the probability would you assign it to happening? (decimal between 0 and 1)

```
survey.data %>% ggplot(aes(x = unlikely)) +
  geom_histogram() + labs(y = "Frequency Count", x = "Probability (Highly Unlikely)")
```



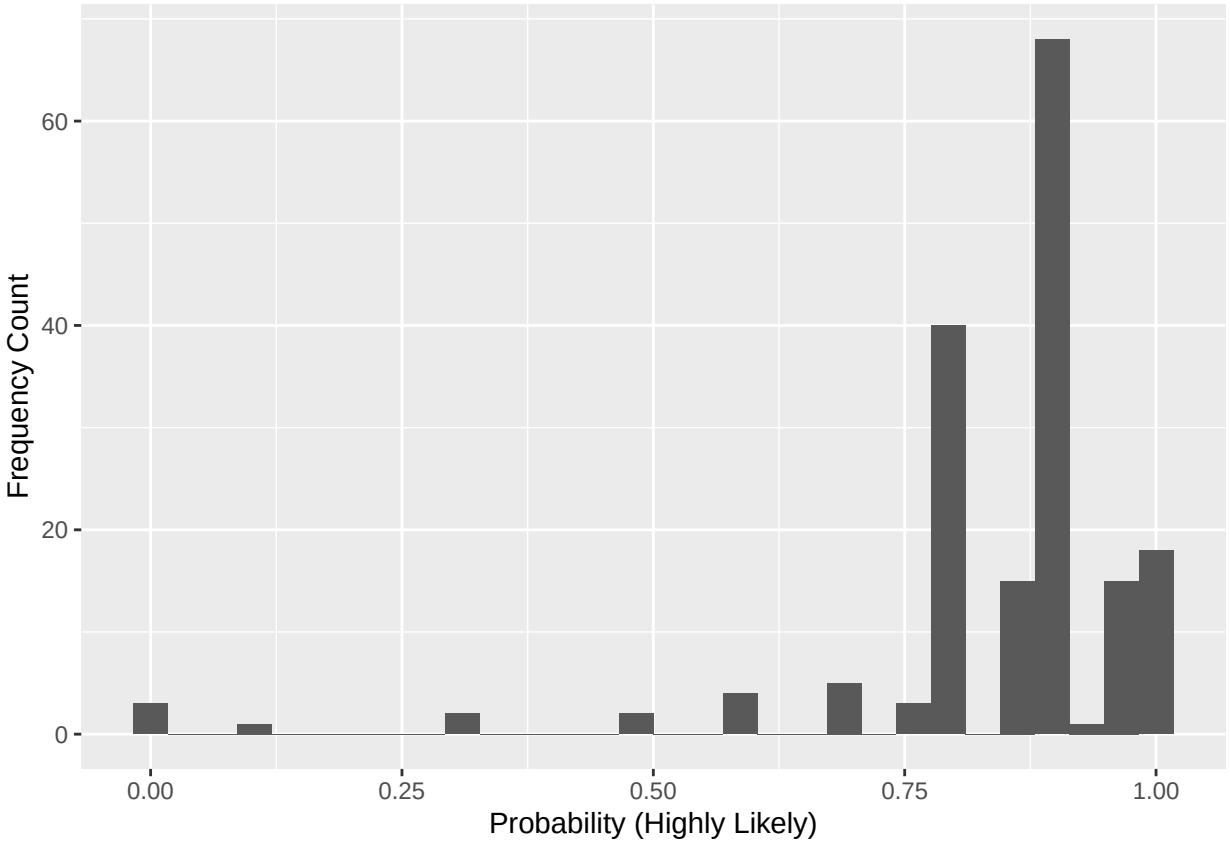
Q15. If you were told it was "Probably" that a natural disaster was about to occur. What is the probability would you assign it to happening? (decimal between 0 and 1)

```
survey.data %>% ggplot(aes(x = probably)) +
  geom_histogram() + labs(y = "Frequency Count", x = "Probability (Probably)")
```



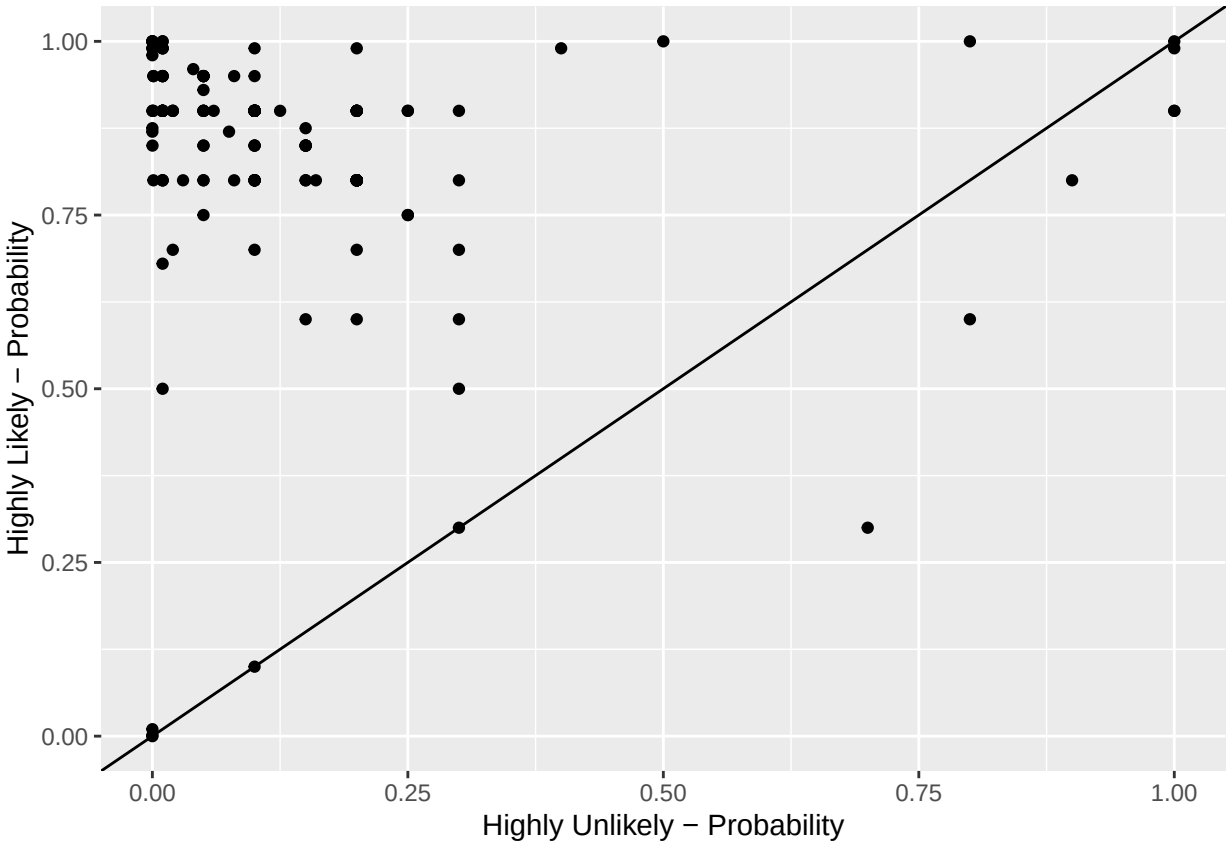
Q16. If you were told it was "Highly Likely" that a natural disaster was about to occur. What is the probability would you assign it to happening? (decimal between 0 and 1)

```
survey.data %>% ggplot(aes(x = likely)) +
  geom_histogram() + labs(y = "Frequency Count", x = "Probability (Highly Likely)")
```



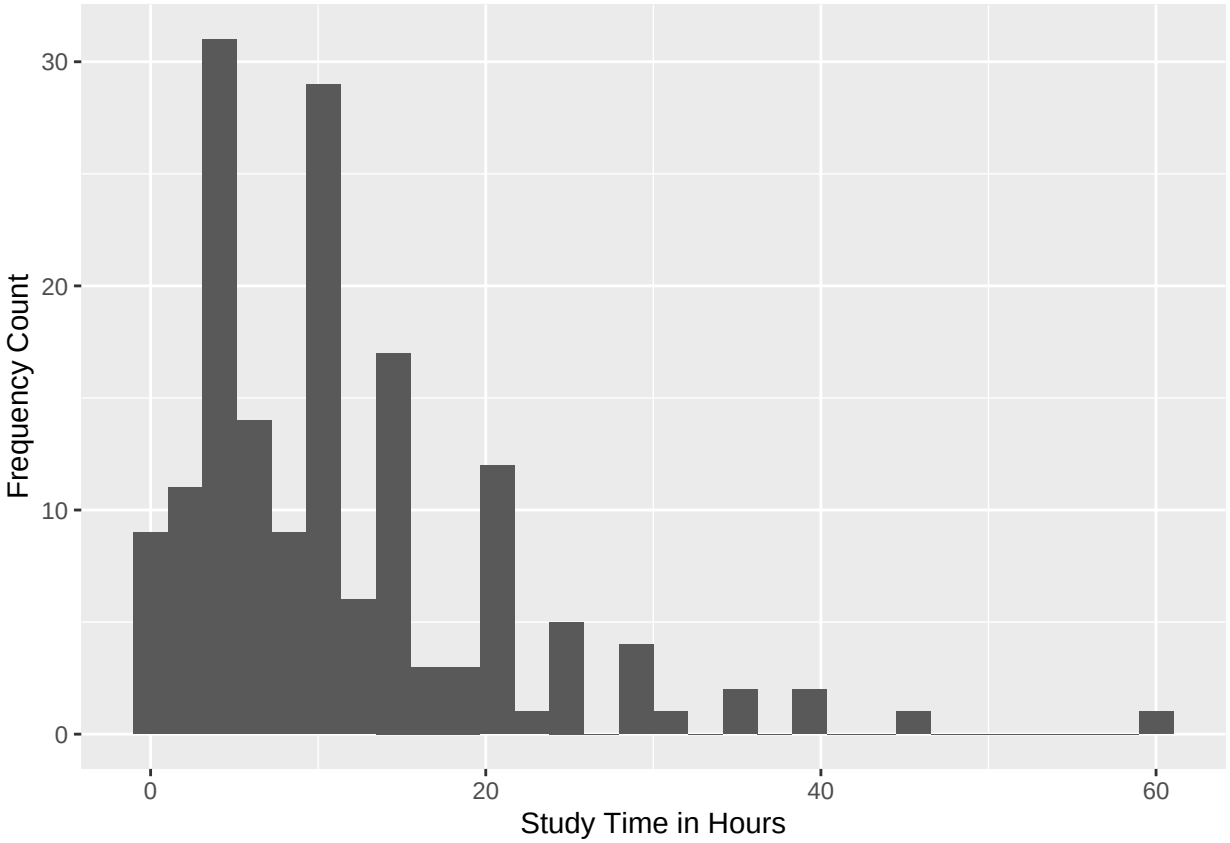
Compare "highly unlikely" and "high unlikely" interpretations.

```
survey.data %>% ggplot(aes(x = unlikely, y = likely)) +
  geom_point() + labs(x = "Highly Unlikely - Probability",
                      y = "Highly Likely - Probability") +
  geom_abline(intercept = 0, slope = 1)
```



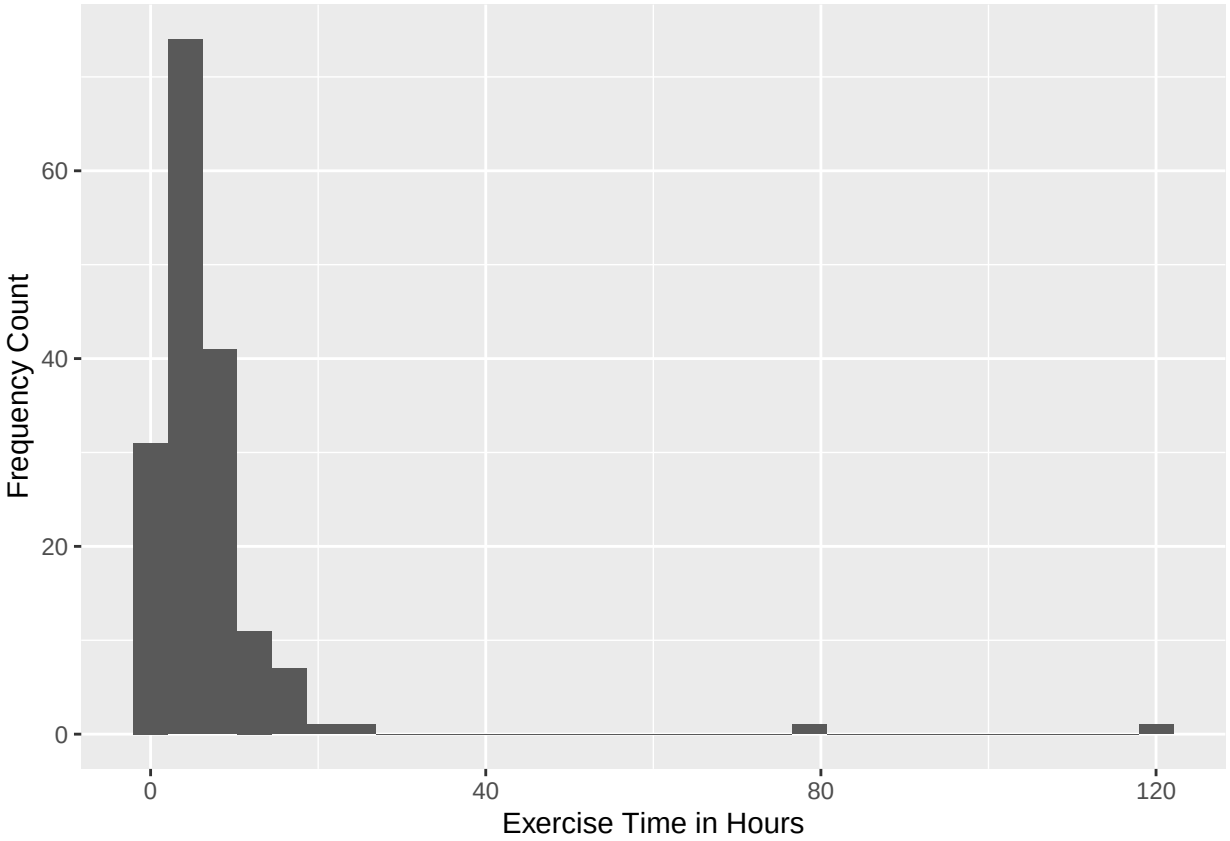
Q17. How many hours in a typical week do you spend studying?

```
survey.data %>% ggplot(aes(x = studying)) +
  geom_histogram() + labs(y = "Frequency Count", x = "Study Time in Hours")
```



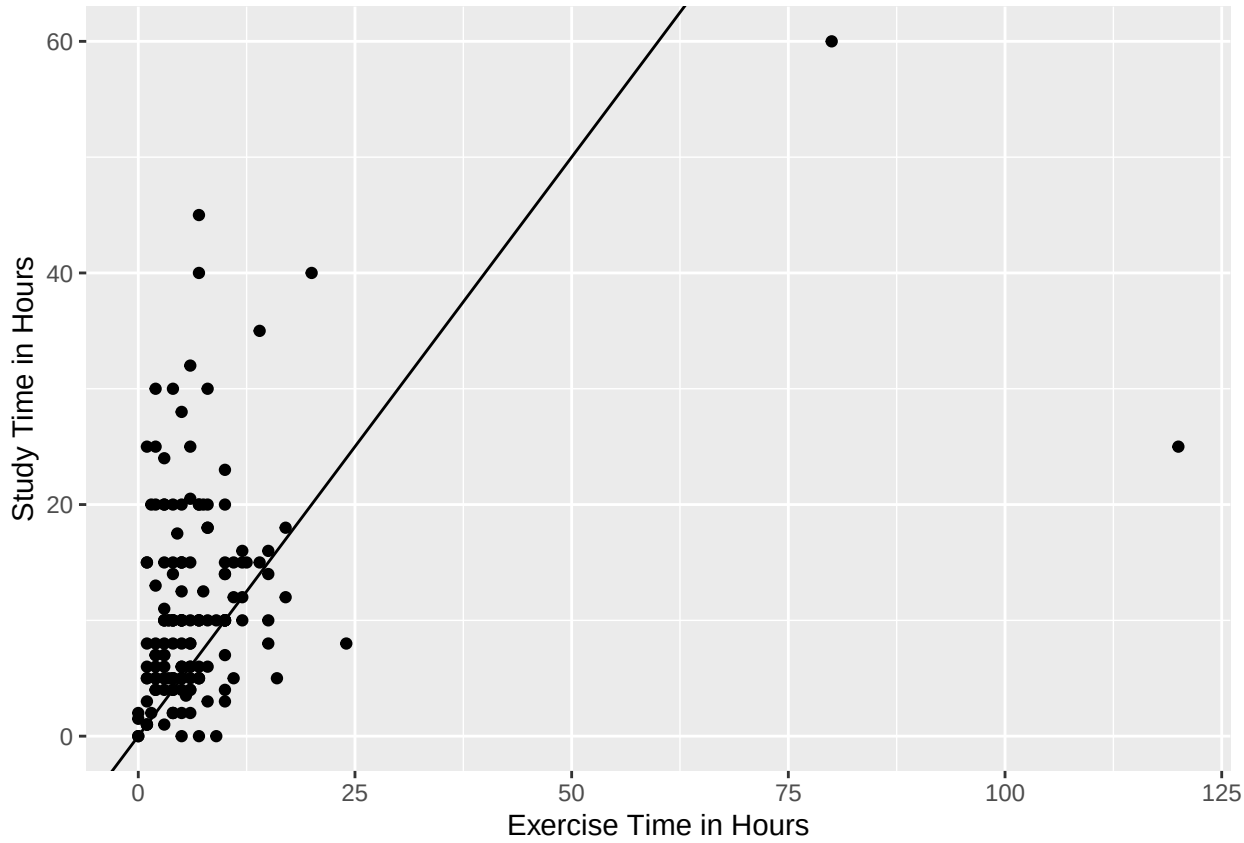
Q18. How many hours in a typical week do you spend exercising?

```
survey.data %>% ggplot(aes(x = exercise)) +
  geom_histogram() + labs(y = "Frequency Count", x = "Exercise Time in Hours")
```



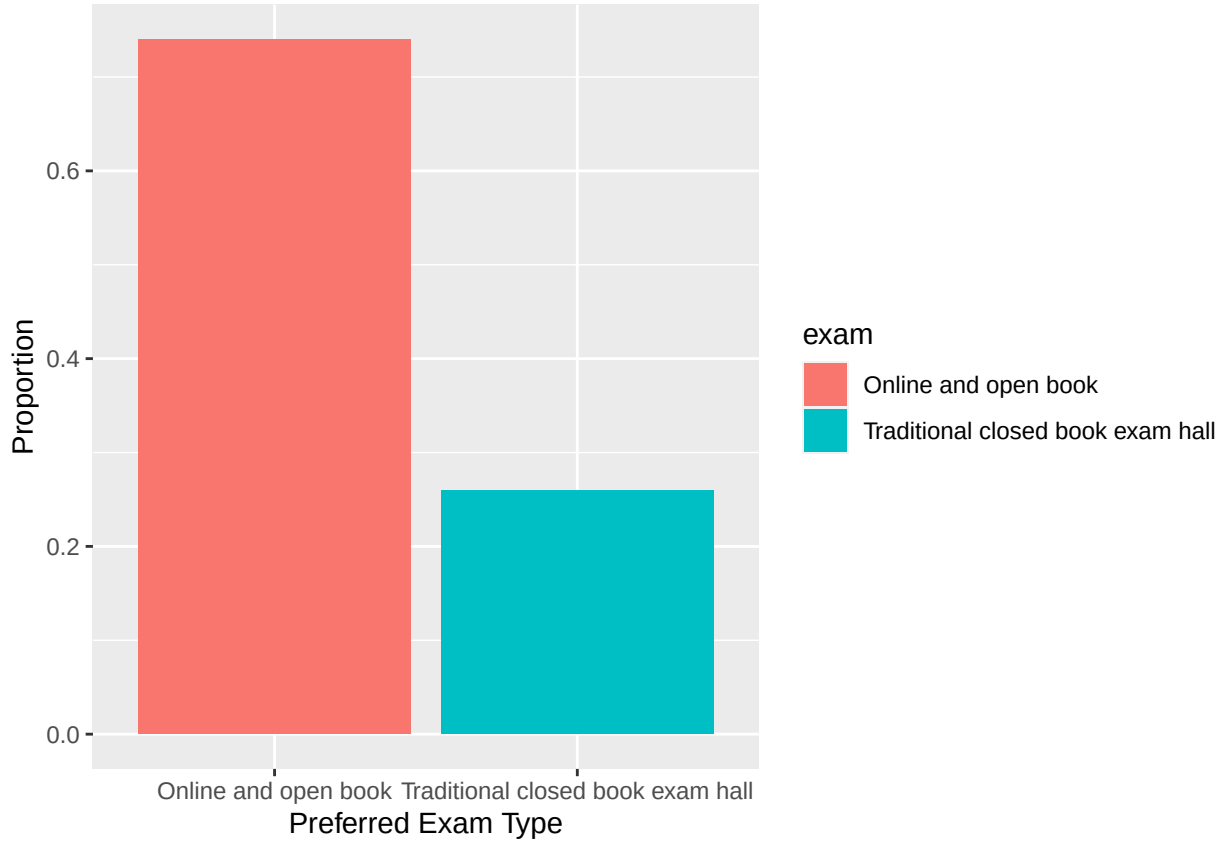
Compare time spent studying and exercising. The line of equality is provided for comparison.

```
survey.data %>% ggplot(aes(x = exercise, y = studying)) +
  geom_point() + labs(x = "Exercise Time in Hours", y = "Study Time in Hours") +
  geom_abline(intercept = 0, slope = 1)
```



Q19. What type of exam do you prefer?

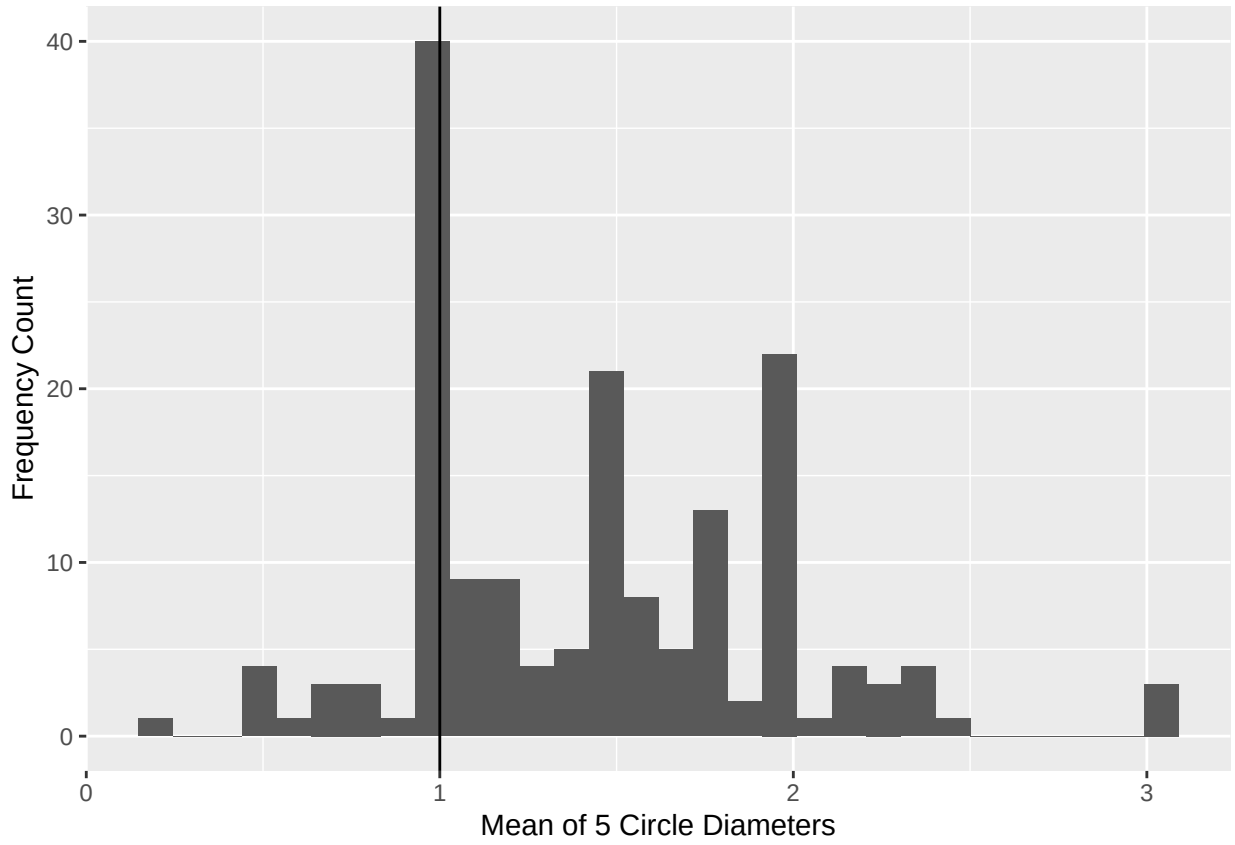
```
survey.data %>% mutate(exam = factor(exam)) %>%
  ggplot(aes(x = exam, y = (..count..) / sum(..count..), fill = exam)) +
  geom_bar() + labs(y = "Proportion", x = "Preferred Exam Type")
```



Q20. Select a representative sample of five circles and calculate their average diameter

The true population average is a diameter of 1. Notice that the central tendency is much higher than this, we'll explain why later!

```
survey.data %>% ggplot(aes(x = circles)) +
  geom_histogram() + labs(y = "Frequency Count", x = "Mean of 5 Circle Diameters") +
  geom_vline(xintercept = 1) # we'll explain why later!
```




```
summary(survey.data$circles)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.156   1.000   1.500   1.460   1.800   3.000      10
```