MT F21 Key

Sheila Weaver

10/17/2021

1. Set-up

```
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse)
surv <- read.csv("SurveyforMTF21.csv", na.strings = '', stringsAsFactors = TRUE)
summary(surv)
```

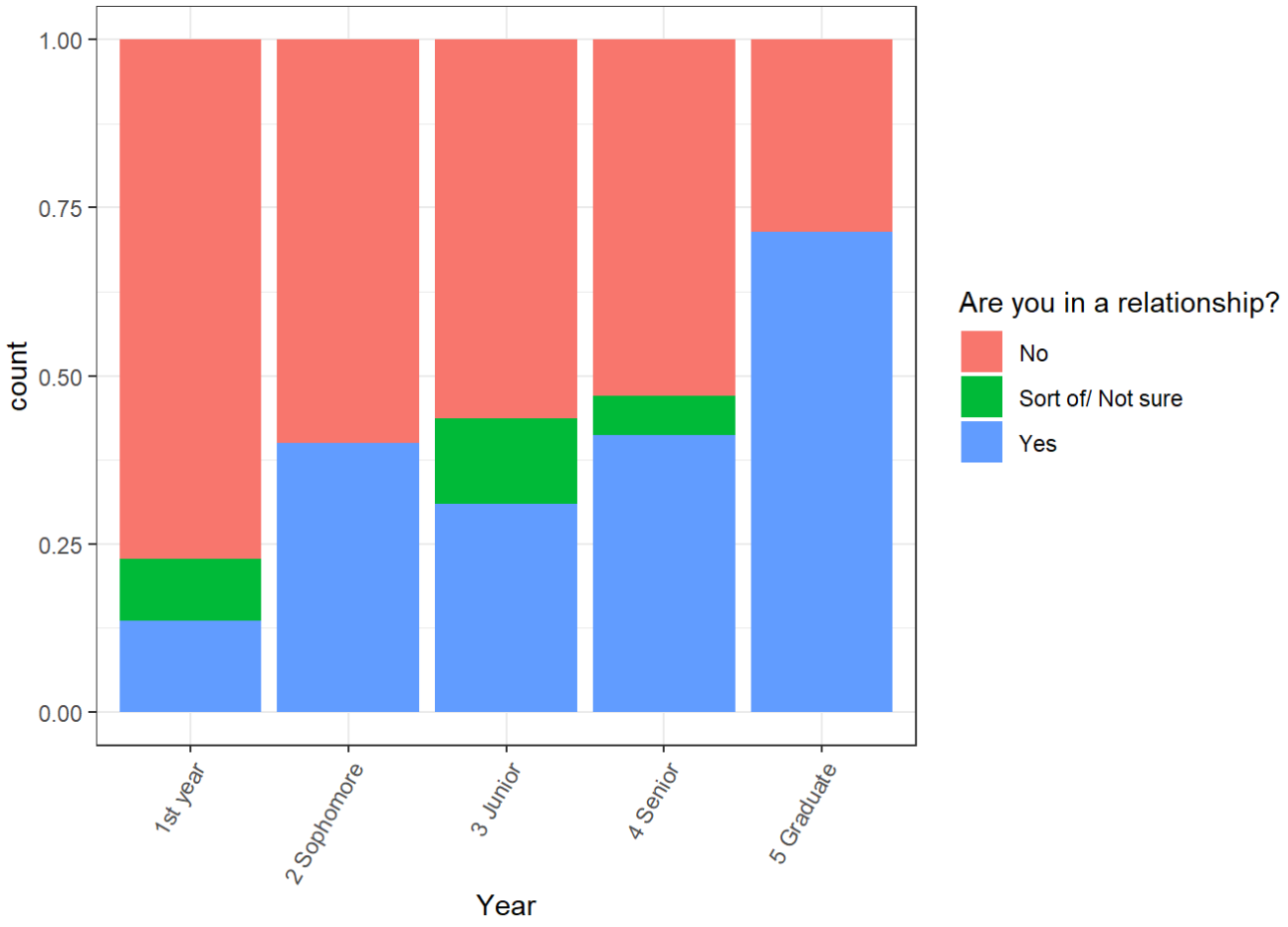
```
##   Response_id                Year       Earnings          Books
##  Min.   :1166955   1st year   :22   Min.   :     0   Min.   :  0.00
##  1st Qu.:1170999   2 Sophomore:25   1st Qu.:  2000   1st Qu.:  1.00
##  Median :1171793   3 Junior   :55   Median :  5000   Median :  3.00
##  Mean   :1171488   4 Senior   :85   Mean   :  6968   Mean   :  5.76
##  3rd Qu.:1172437   5 Graduate : 7   3rd Qu.:  8000   3rd Qu.:  5.00
##  Max.   :1174326   NA's       : 2   Max.   :120000   Max.   :100.00
##                                      NA's   :2
##              Relationship     Sleep           PhoneTime       Breakfast
##  No               :112   Min.   : 4.000   Min.   :0.500   Min.   :0.000
##  Sort of/ Not sure: 14   1st Qu.: 7.000   1st Qu.:3.000   1st Qu.:3.000
##  Yes              : 70   Median : 7.500   Median :3.750   Median :5.000
##                          Mean   : 7.365   Mean   :3.732   Mean   :4.898
##                          3rd Qu.: 8.000   3rd Qu.:4.125   3rd Qu.:7.000
##                          Max.   :11.000   Max.   :9.000   Max.   :7.000
##
##                          Triangle        GPA
##  Sleep and Good Grades       : 71   Min.   :1.900
##  Sleep and Social Life       : 12   1st Qu.:3.200
##  Social Life and Good Grades:112   Median :3.555
##  NA's                        :  1   Mean   :3.449
##                                     3rd Qu.:3.800
##                                     Max.   :4.000
##                                     NA's   :40
```

2. Bar Graph

The trend that we can observe here, is that as university students as they get older (pass from freshmen to sophomore, etc) they are more likely to be in a relationship. This can be seen in the graph by looking at how the ‘pink’ portion of each bar gets smaller from left to right, while the ‘blue’ part gets larger. An explanation for this trend is that when once gets older, the desire to settle down grows larger, as well as the fact that you get to know more people (and also get to know them better), the more time you are in college.

```
# data frame to use in the bar graph with the NAs removed from Year
surv2 <- surv %>% filter(!is.na(Year))

# Bar graph of Relationship by Year
ggplot(data = surv2,
       mapping = aes(x = Year, fill = Relationship)) +
  geom_bar(position = 'fill') +
  theme_bw() +
  theme(axis.text.x = element_text(hjust=1, angle = 60)) +
  labs(fill = "Are you in a relationship?")
```

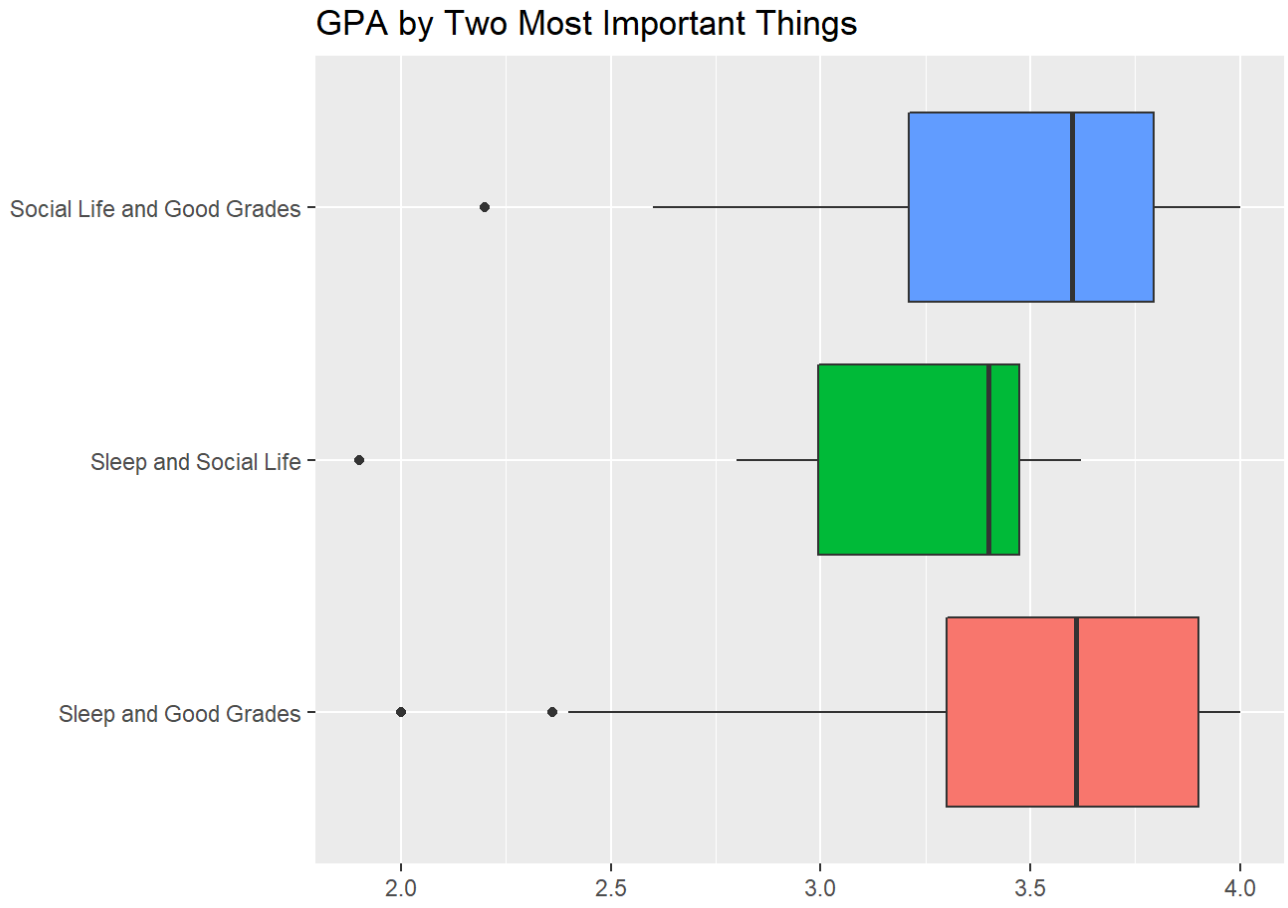


3. Boxplots

In this graph we can see that the obvious happens: people who prioritize “Sleep and Social Life” get worse grades. However, what is very interesting is that the there is not a large difference regarding the median grade of people who prioritize “Social Life and Good Grades” to those who prefer “Sleep and Good Grades”. Nevertheless, the people who obtain the maximum grades overall are those who prefer to sleep over having a good social life. This makes a lot of sense because sleeping is essential to correct functioning of the brain.

```
# filtering of data through dpylr to remove NA in Triangle
surv3 <- surv %>% filter(!is.na(Triangle))

#  boxplot graph of GPA by Triangle
ggplot( data = surv3,
        mapping = aes( y = Triangle, x = GPA , fill = Triangle)) +
  geom_boxplot() +
  labs(title="GPA by Two Most Important Things", x = "", y = "" ) +
  guides(fill=FALSE) +
  theme_gray()
```



4. Stats of Earnings by Year

```
surv %>% # data to be used
  filter(!is.na(Year))%>% # removing NAs from the Year variable
  group_by(Year) %>%  # grouping the data into groups according to Year (freshmen,
sophomore, etc)
  summarise(mean_earnings = mean(Earnings), # get mean
            median_earnings = median(Earnings), # get median
            n_students = n()) # get number of students
```

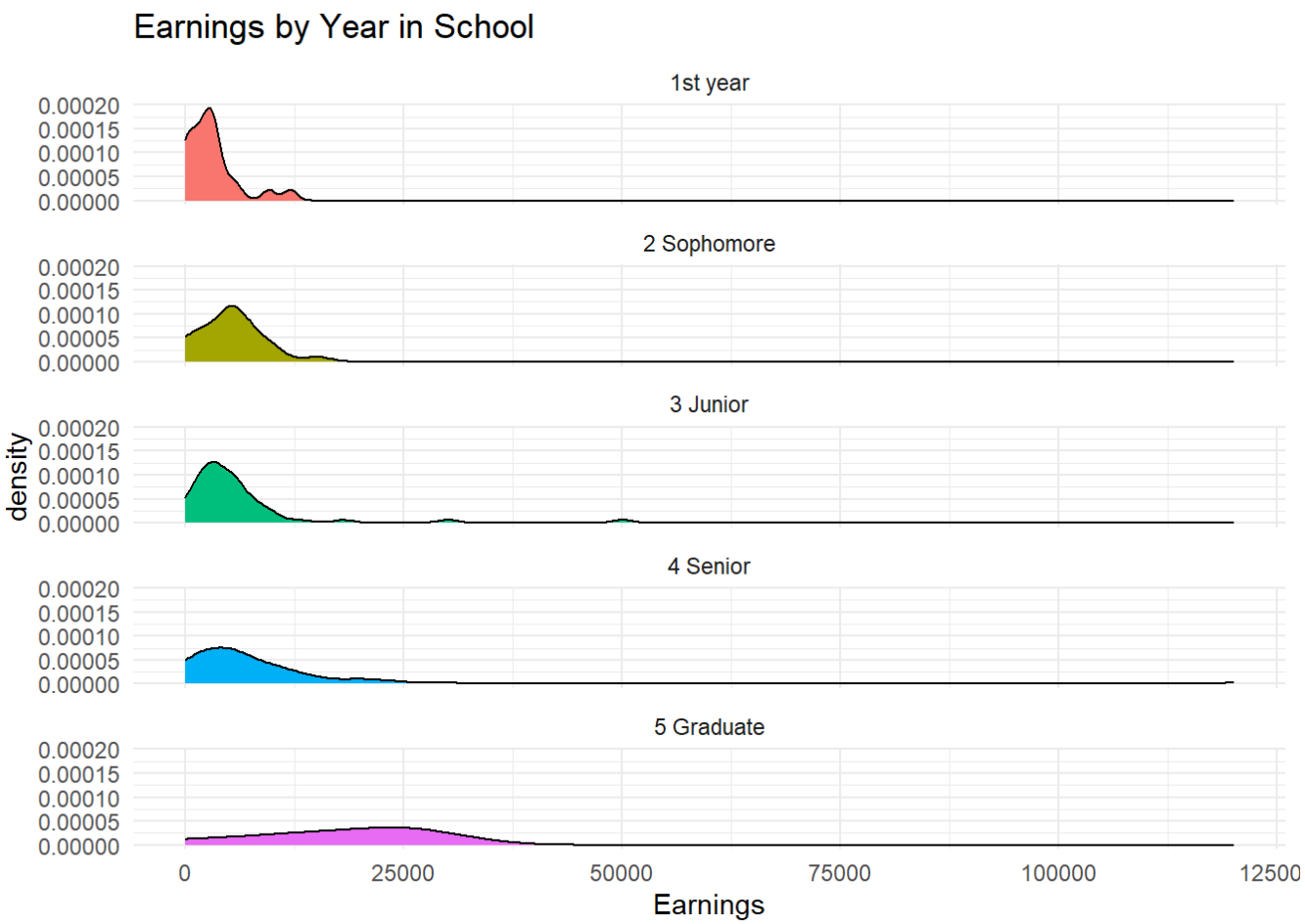
```
## # A tibble: 5 x 4
##   Year        mean_earnings median_earnings n_students
##   <fct>               <dbl>           <dbl>      <int>
## 1 1st year            3018.            2750         22
## 2 2 Sophomore         5210             5000         25
## 3 3 Junior            5868.            4000         55
## 4 4 Senior            8345.            5307         85
## 5 5 Graduate         17571.           20000          7
```

5. Density plots of Earnings by Year

I do believe that the Earnings and Year variables are related because as we can see in both the graph and the summary variables previously shown, is that as the student progresses academically, the salary they received increases. This is specially true between Seniors and Graduates. Apart from this, it really caught my eye how the group with the least of people working (least density of earnings) are seniors, this could possibly be because the courseload increases during each academic year. Another very noticeable trend is that graduate students tend to earn a lot more than undergraduates.

```
# surv4 is a data frame to filter out the NAs from the Earnings
surv4 <- surv %>% filter(!is.na(Earnings))

# density graph of Year by Earnings
ggplot( data = surv4,
        mapping = aes( x = Earnings, fill=Year )) +
  geom_density() +
  facet_wrap( ~ Year,  ncol = 1) +
  theme_minimal() +
  guides(fill='none') +
  labs(title="Earnings by Year in School")
```



6. Top 8 and Bottom 8 GPAs

The main differences I could notice between the two groups are that the ones with highter GPA tend to eat on average more times a week breakfast and read more books for pleasure.

```
# Creating the data frame to use in both cases
surv4 <- surv %>% # naming of the data frame and saying what data to use
  filter( !is.na(Year) ,!is.na( GPA) ,!is.na(PhoneTime) ,!is.na(Sleep) ,!is.na( Bre
akfast) ,!is.na(Books)) %>% # removal of NAs
  select( Year, GPA, PhoneTime, Sleep, Breakfast, Books) # selecting columns to app
ear

# The 8 students with highest GPA
surv4 %>% # data
  arrange(desc(GPA)) %>% # telling dplyr to arrange them in descending order accord
ing to GPA
  head(8) # selecting the first 8 entries
```

```
##          Year  GPA PhoneTime Sleep Breakfast Books
## 1    3 Junior 4.00         2   8.5         7     4
## 2    3 Junior 4.00         2   6.5         5     5
## 3    3 Junior 4.00         4   5.0         7     8
## 4 2 Sophomore 4.00         4   8.0         4     5
## 5  5 Graduate 4.00         3   8.0         7    15
## 6    4 Senior 3.98         2   8.0         3    10
## 7    4 Senior 3.98         4   7.5         7     5
## 8 2 Sophomore 3.97         4   8.0         5     2
```

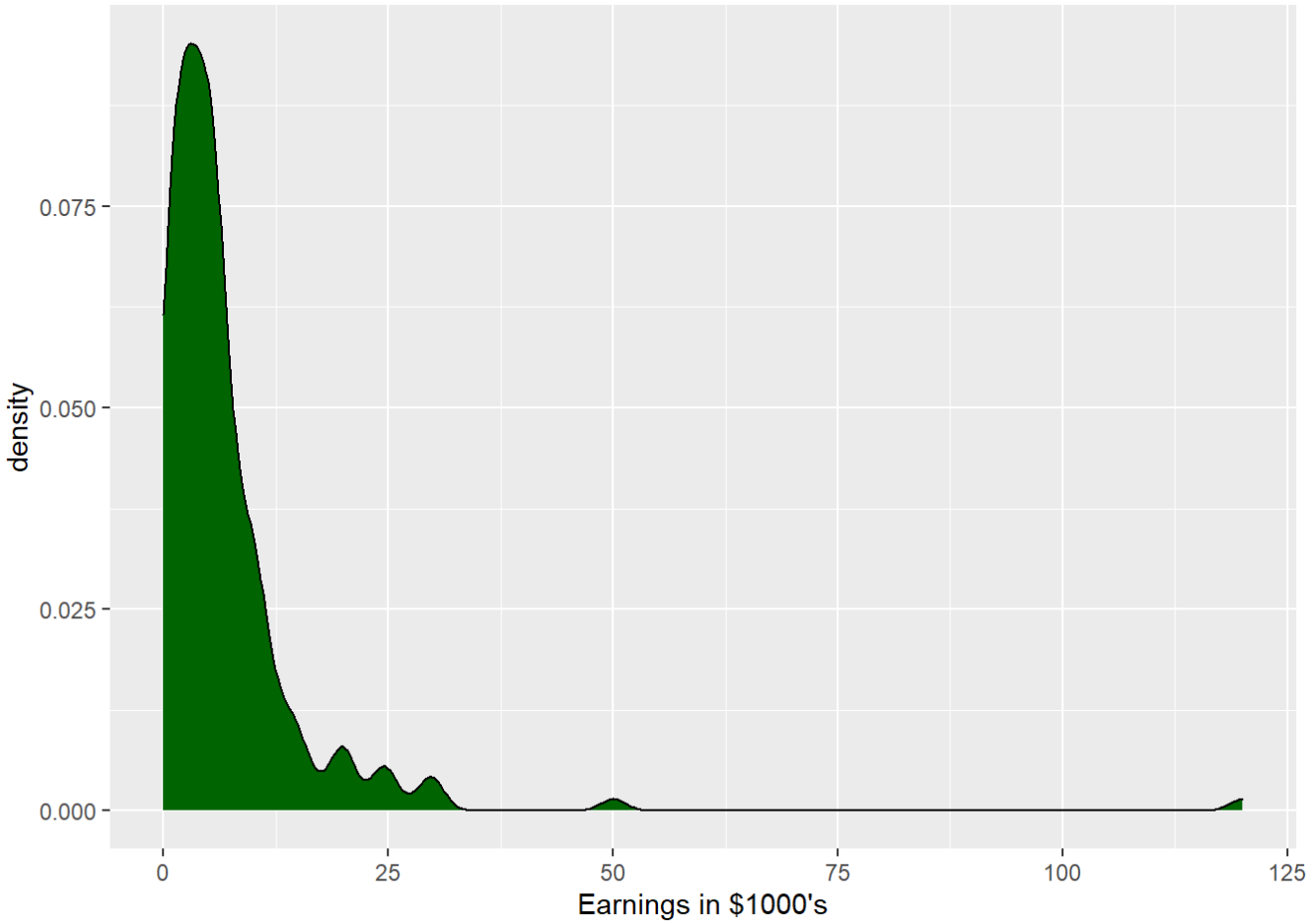
```
#  The 8 students with the lowest GPA
surv4 %>% # data
  arrange(desc(GPA)) %>% # telling dplyr to arrange them in descending order accord
ing to GPA
  tail(8) # selecting the last 8 entries
```

```
##         Year  GPA PhoneTime Sleep Breakfast Books
## 147 4 Senior 2.67       2.0   8.0         7     4
## 148 4 Senior 2.60       2.0   5.5         3     1
## 149 3 Junior 2.50       3.5   9.0         0     3
## 150 4 Senior 2.40       4.0   8.0         3     7
## 151 4 Senior 2.36       2.0   7.0         3     4
## 152 4 Senior 2.20       6.0   4.0         2     3
## 153 4 Senior 2.00       1.0   8.0         7     0
## 154 4 Senior 1.90       4.0   8.0         6     4
```

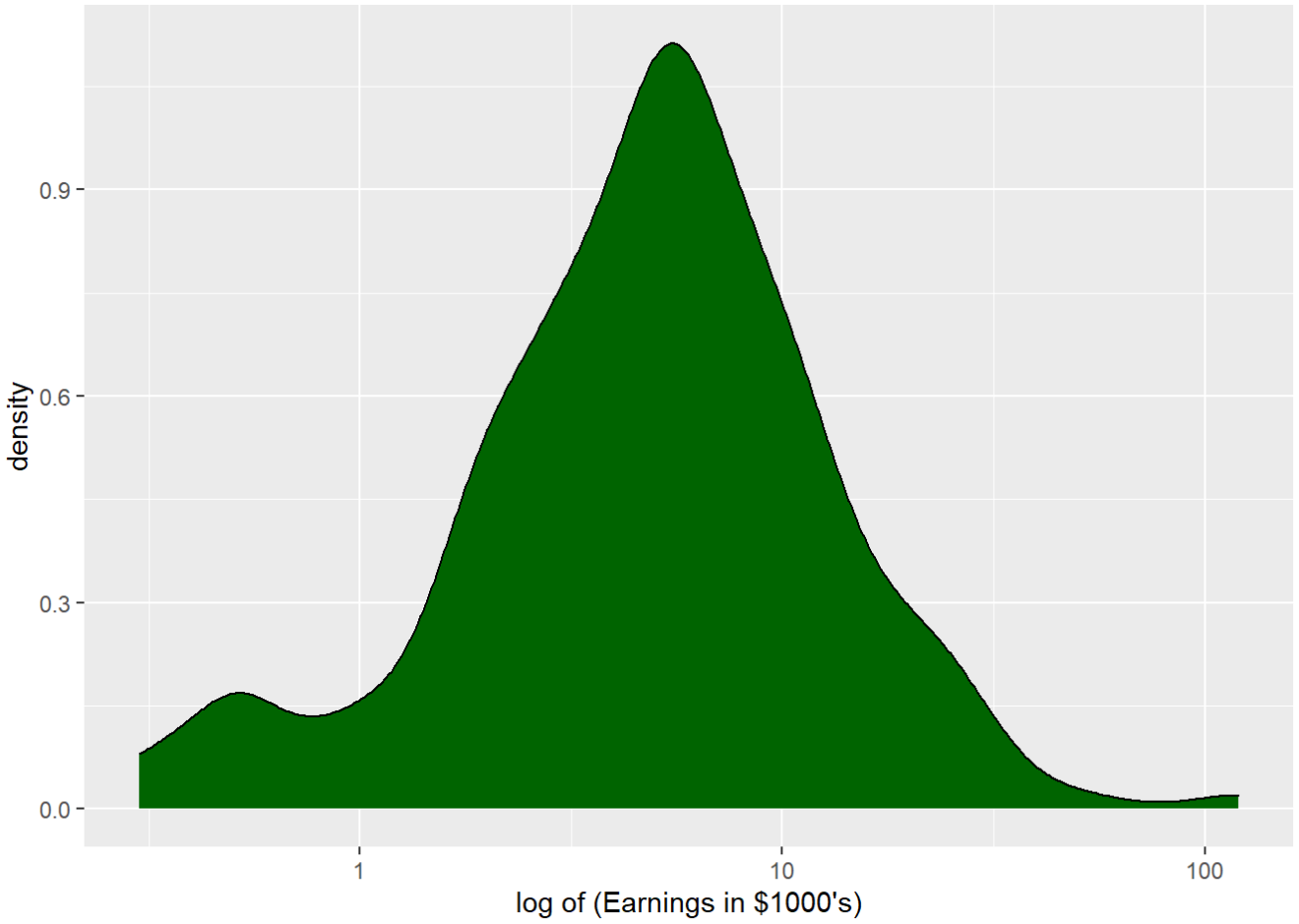
7. Density plots of Earnings

```
# Creating data frame with the surveys' data to
surv7<- surv %>% # naming of dataframe and data to use
  filter(!is.na(Earnings))%>% # removing NAs from Earnings variable
  mutate(earnings_1000 = (Earnings / 1000)) # creating new variable where Earning's
appears in 1000s of dollars

# density plot of Earnings in thousands
ggplot(data = surv7,
       mapping = aes( x= earnings_1000) ) +
  geom_density(fill="darkgreen") +
  labs(x = "Earnings in $1000's")
```




```
# density plot of log of earnings in 1000s
ggplot(data = surv7,
       mapping = aes( x= earnings_1000) ) +
  geom_density(fill="darkgreen") +
  scale_x_log10() +
  labs(x = "log of (Earnings in $1000's)")
```



Statement of Academic Honesty:

I confirm that I, Lucía Carrera, did not consult with any other person while doing this test, either in person, email, texting, etc. I did not look at any other person's work, or show my work to others.