Project 1: Sarah Lee

Code

1. Introduction

a. Set up

Hide

```
# Load packages
library(tidyverse)
library(tidytext)
library(textdata)
```

b. Datasets

https://www.kaggle.com/datasets/lislejoem/us-minimum-wage-by-state-from-1968-to-2017 (https://www.kaggle.com/datasets/lislejoem/us-minimum-wage-by-state-from-1968-to-2017)

https://www.kaggle.com/datasets/tjkyner/bachelor-degree-majors-by-age-sex-and-state (https://www.kaggle.com/datasets/tjkyner/bachelor-degree-majors-by-age-sex-and-state)

Hide

```
wage <- read_csv("Downloads/Minimum Wage Data.csv")
```

```
Rows: 2862 Columns: 15── Column specification ────────────────────────────────
────────────────────────────────────
Delimiter: ","
chr  (3): State, Department.Of.Labor.Uncleaned.Data, Footnote
dbl (12): Year, State.Minimum.Wage, State.Minimum.Wage.2020.Dollars, Federal.Minimum.
Wage, Feder...
ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Hide

```
degree <- read_csv("Downloads/Bachelor_Degree_Majors.csv")
```

```
Rows: 612 Columns: 9── Column specification ────────────────────────────────────────
────────────────────────────────────

Delimiter: ","
chr (3): State, Sex, AgeGroup
num (6): Totaldegholders, Science and Engineering, Science and Engineering Related Fi
elds, Busin...
ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Hide

```
wage
```

Y... State <dbl><chr>	State.Minimum.Wage <dbl>	State.Minimum.Wage.2020.Dollars <dbl>
1968 Alabama	0.00000	0.00
1968 Alaska	2.10000	15.61
1968 Arizona	0.46800	3.48
1968 Arkansas	0.15625	1.16
1968 California	1.65000	12.26
1968 Colorado	1.00000	7.43
1968 Connecticut	1.40000	10.41
1968 Delaware	1.25000	9.29
1968 District of Columbia	1.25000	9.29
1968 Florida	0.00000	0.00

1-10 of 2,862 rows | 1-4 of 15 columns Previous **1** 2 3 4 5 6 … 100 Next

Hide

```
degree
```

State <chr>	Sex <chr>	AgeGroup <chr>	Totaldegholders <dbl>	Science and Engineering <dbl>
Alabama	Total	25 and older	885357	263555
Alabama	Total	25 to 39	268924	90736

Alabama	Total	40 to 64	418480	115762
Alabama	Total	65 and older	197953	57057
Alabama	Male	25 and older	405618	159366
Alabama	Male	25 to 39	117794	50188
Alabama	Male	40 to 64	184328	67977
Alabama	Male	65 and older	103496	41201
Alabama	Female	25 and older	479739	104189
Alabama	Female	25 to 39	151130	40548

1-10 of 612 rows | 1-5 of 9 columns Previous **1** 2 3 4 5 6 ... 62 Next

Hide

```
NA
```

Wage displays data of minimum wage by each state from 1968 to 2020 while degree displays data of bachelor's degree majors by age, sex, and state. Both of these data sets were extracted from Kaggle which is linked above. I found the relationship between major types and minimum wages to be intriguing.In 'wage', there are 15 variables exploring types of minimum wages (numerical variables), like minimum wage in 2020 dollars as the value money changes throughout time/inflation, federal minimum wage, or effective minimum wage. In 'degree', there are 9 variables exploring the 5 types of degree, age (in 3 age groups, making it a categorical variable), sex, and state. I plan to joining these two data sets through the 'state' variable which will ultimately compare the degree type and minimum wage.

c. Define a research question

Does the type of bachelor's degree major influence minimum wage in the US?

2,3,4. Make our data analyzable

Tidying, joining, and wrangling

Hide

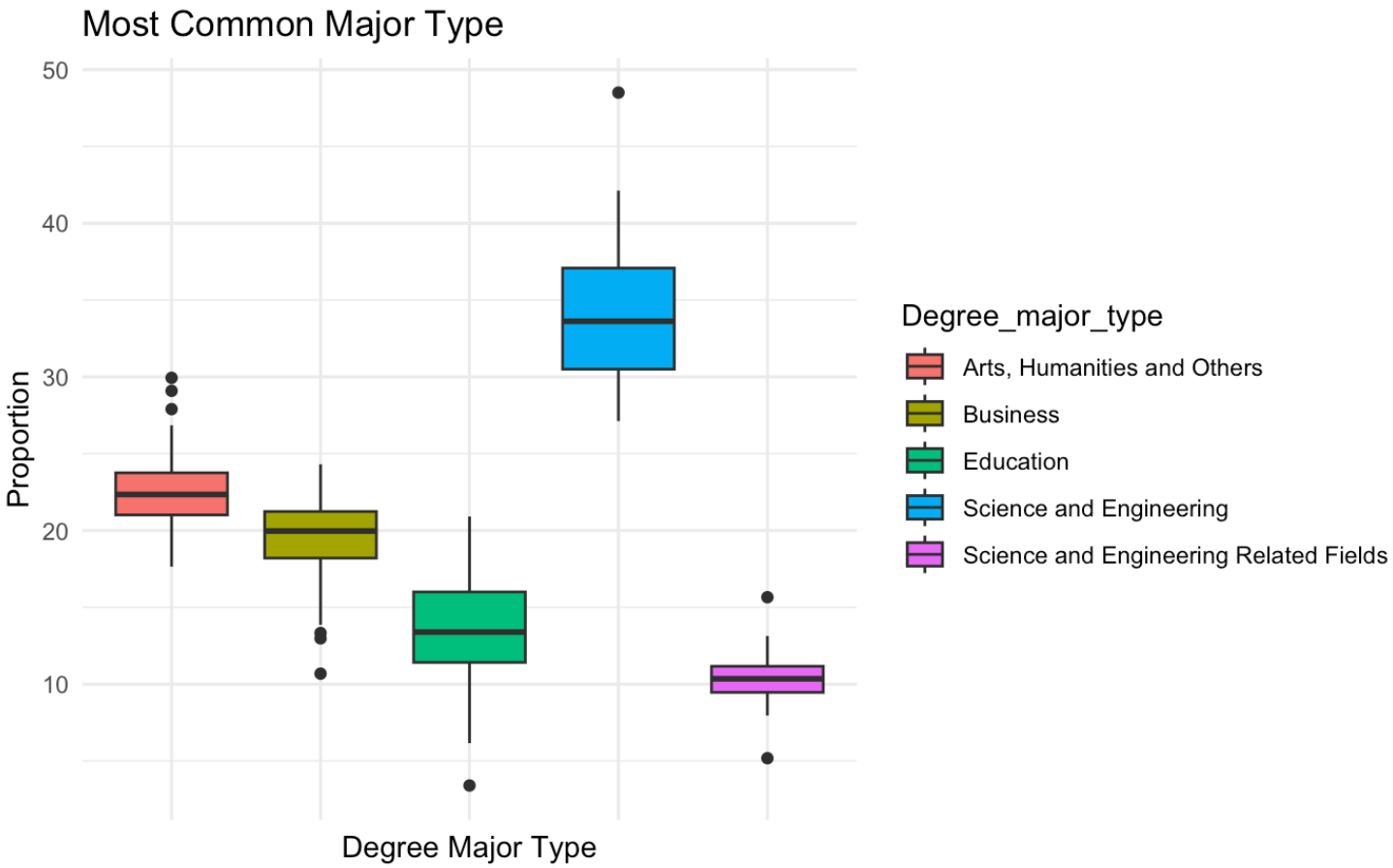
```
wage1=wage%>%
  filter(Year=="2019") #only keep info in 2019 as 'degree' data set was formed in 201
9

degree1=pivot_longer(degree,
             cols = c('Science and Engineering':'Arts, Humanities and Others'),
             names_to = "Degree_major_type",
             values_to = "Count")%>% #organized long data set into each degree major
type
  mutate(prop=(Count/Totaldegholders)*100)%>% #proportion/distribution of major types
within each state to see which major type was the most or least common for each state
  filter(Sex=="Total", AgeGroup== "25 and older") #this filter considers most of the
data set as we are not analyzing by sex or age

joined = left_join(degree1,wage1,by="State") #joining two data sets by the common var
iable "State"


wage1 %>%
  summarize(n_wage = n_distinct(State.Minimum.Wage)) #27 observations
degree1%>%
  summarize(n_degree= n_distinct(prop)) #255 observations
joined %>%
  summarize(n_joined= n_distinct(prop)) #255 observations, since this is a left joing
of wage1 into degree1, the final joined observations equal degree1's observations

myjoined=joined%>%
  select(1:9) #eliminated all unnecessary variables to display only necessary informa
tion
  group_by(Totaldegholders)%>% #grouping by the total degree holders so that we see w
hich major was the most common in that state
  arrange(desc(prop)) #displays which major was the most common- Science and Engineer
ing
```

5. Visualize and analyze our data

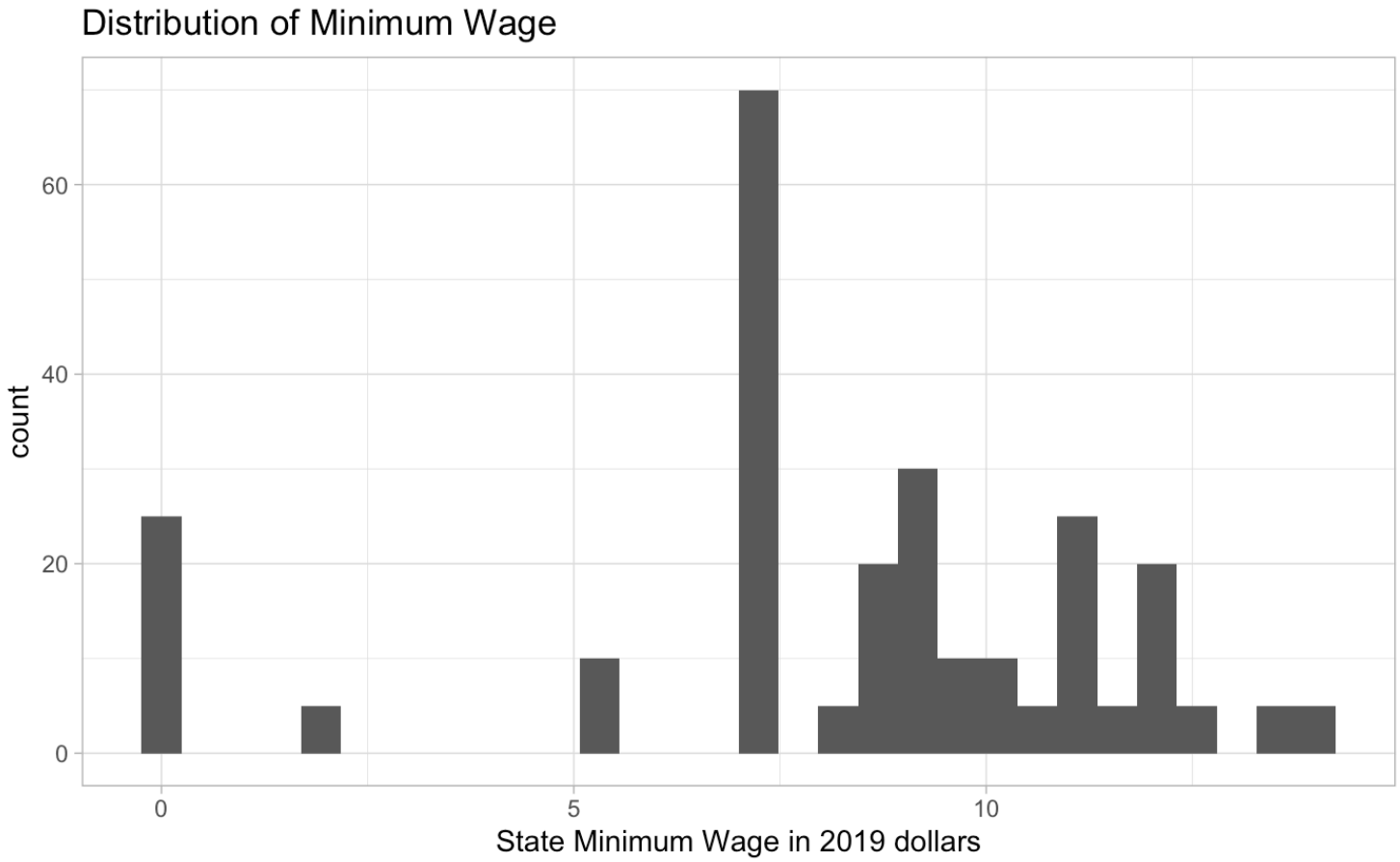
Taking the information from above, I have created three visualizations to help further the understanding of my analysis.

Hide

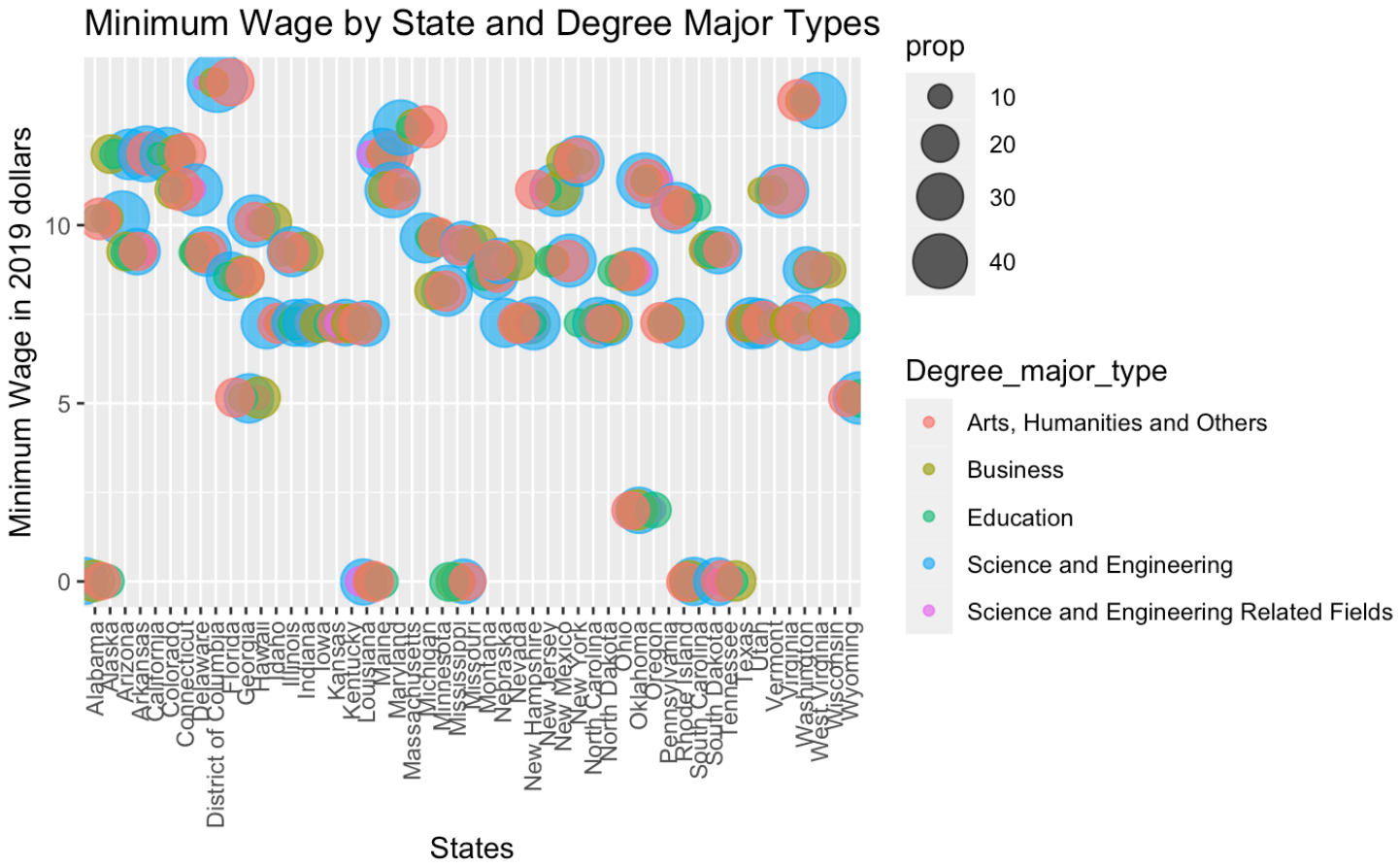
```
myjoined%>%
  ggplot(aes(x=Degree_major_type,y=prop,fill=Degree_major_type))+
  geom_boxplot()+
  labs(title="Most Common Major Type",x="Degree Major Type",y="Proportion")+
  theme_minimal()+
  theme(axis.text.x=element_blank())
```



This box plot displays that Science and Engineering degree majors were the most common among the states as it has the highest mean and maximum values.



The histogram above explores the distribution of minimum wage of the states in the US. We are able to see that the most common minimum wage among the states was approximately a little less than $7.5.



This final plot shows the minimum wage for each state. The sizes of data points are dependent of 'prop' or the proportion of the specific major type within each state since I divided the count of the specific major by the total number of bachelor degree holders in each state. This process was done so that the size of the population of each state didn't influence the counts of my data set. According to the dataset the sizes of the points got bigger as the amount of minimum wage increased. Blue or Science and Engineering was mostly seen as the amount of minimum wage increased as well.

6. Discussion

Overall, through visualizations and analysis, I was able to conclude that Bachelor's degree majors type does influence minimum wage. Specifically, degree major of Science and Engineering plays a big role in the higher minimum wage according to graphs 1 and 3. In graph 3, we see that the proportions become much bigger for Science and Engineering as we look at higher minimum wages of the states. In this analysis, I eliminated many variables from the two original data sets to observe specific variables but I think it'll be important to study the data more deeply considering other factors like inflation, gender, age, and time.