What Major Should You Choose? Comparison of Median Salary Based on Career Status, Salary Growth, and Unemployment Rates

Contents

1. Introduction

One of the most important life decisions an individual makes occurs when he/she is just a high school student: what major to pursue. What major and degree one obtains is critical for job after college, subsequently setting their career trajectory. Therefore, we extracted and visualized data to see which college degrees offer the most pay, so that students are able to make a more data-driven decision.

First, we wanted to visualize the distribution of the median starting and mid-career salaries of each major/degree category. Then, we plotted the top 5 majors of each major category based on the median salary. We were also interested in looking at salary growth as one progressed from starting to mid-career salary. So, we visualized the percent change in median starting salary to mid-career salary for each of the major categories, and plotted the top 20 majors by salary percent change.

Data for salary distributions and salary growth were obtained from the FiveThirtyEight data repository (Link: fivethirtyeight/data), which contained aggregated data from 2010 - 2012. Each unique row contains: the major, major category, total number of people graduated with the major, sample size, number employed, number unemployed, unemployment rate, median salary, P25 salary, and P75 salary. The major and major category are categorical variables, while the other variables are numeric. This dataset contained data of the variables mentioned above for both recent graduates/starting career (ages < 28) and graduates/mid-career graduates (ages 28+). The datasets were joined using the major as the key.

In addition, we wanted to visualize how median salary changed in the past decade, and how unemployment rates compare to the national (U.S.) unemployment rate for each major category in 2021. In order to obtain the information needed, we extracted the ACS 1-Year PUMS data from the United States Census Bureau (Link: PUMS data) from 2009 - 2021. The data we extracted were major, number employed, number unemployed, and median salary. Like previously done, the datasets were joined using major as the key. The national unemployment rate data of 2021 was obtained from the U.S. Bureau of Labor Statistics (Link: Unemployment Data), and the 2021 US unemployment rate was calculated by taking the average unemployment rate of each month in 2021.

This report aims to address two research questions: 1) Which college degree should you pursue for a career with a high starting and mid-career salary 2) Which college degree shows the most promising growth rate from starting to mid-career salary as well as in respect to time?

a. Set up

b. Dataset Creation

```
# create data frame
grad_students_data <- read.csv("~/Desktop/SDS322E/Project/project data/grad-students.csv")
recent_grads_data <- read.csv("~/Desktop/SDS322E/Project/project data/recent-grads.csv")

# remove irrelevant variables
grads <- grad_students_data %>%
  select(-"Grad_full_time_year_round", -"Nongrad_total",
         - "Nongrad_employed", - "Nongrad_full_time_year_round",
         - "Nongrad_unemployed", - "Nongrad_unemployment_rate",
         - "Nongrad_median", - "Nongrad_P25",
         - "Nongrad_P75", - "Grad_share",
         - "Grad_premium", - "index", - "Major_code") %>%
  arrange(desc(Grad_median)) %>%
  mutate(Rank = seq(1:173),
# add categorical variable
         Status = "Grad") %>%
```

```
  relocate(Rank, .before = Major)

# remove irrelevant variables
recent_grads <- recent_grads_data %>%
  select(- Men, - Women, - ShareWomen, - Employed,
         - Full_time, - Part_time, - College_jobs,
         - Non_college_jobs, -Low_wage_jobs, -index, - Rank,
         - Major_code) %>%
  # add categorical variable
  mutate(Status = "Recent_Grad")

# rename variables
n_recent_grads <- recent_grads %>%
  rename(Employed = Full_time_year_round)

# rename variables
n_grads <- grads %>%
  select(-"Rank") %>%
  rename(Total = Grad_total, Sample_size = Grad_sample_size, Employed = Grad_employed,
         Unemployed = Grad_unemployed, Unemployment_rate = Grad_unemployment_rate,
         Median = Grad_median, P25th = Grad_P25, P75th = Grad_P75)

# merge datasets using major as the key
combined_data <- left_join(grads, recent_grads, by = "Major")

# reformat merged data dataset to make it tidy
new_combined_data <- rbind(n_grads, n_recent_grads)
```

Data merging: Before joining the datasets, each dataset contained 173 rows. The 'recent-grads' dataset contained 22 columns, and the 'grad-students' dataset contained 23 rows. Relevant variables in common were the major, major category, and median salaries (in addition to other variables listed in the introduction). Before merging, we added a 'Status' variable to indicate whether a given row represents 'recent-grads' data or 'grad-students' data. After merging with 'left_join' by using major as the key (as the majors were identical between the two datasets), the combined dataset contained 173 observations with 23 variables after irrelevant variables were removed. We also created a second merged dataset (which was used for much of the data analysis below) using 'rbind' to stack the two datasets on top of each other rather than next to each other as done with 'left_join' in order to make the data tidy. This new merged dataset (using 'rbind') contained 346 rows with 11 columns.

This means that in total, 34 variables were dropped due to either merging or intentional removal. Variables that were dropped by merging include major category and median salary (as they were overlaps between the two datasets). Variables that were dropped by intentional removal include rank and percentage of women vs men because those variables were not overlaps and/or not relevant for our research questions.

2. Data Analysis

a. Overall Data

I) Average Starting and Mid-Career Median Salary

```r
# Distribution of average median salary of each major category by age status
new_combined_data %>%
  ggplot(aes(x = reorder(Major_category, -Median), y = Median, fill = Status)) +
  geom_bar(stat = "summary", fun = "mean", position = "dodge", width = 0.7) +
  scale_y_continuous(breaks = seq(20000, 100000, 10000)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_fill_discrete(labels = c("Graduate", "Recent Graduate")) +
  labs(title = "Average Median Salary Based on Major Category and Career Status",
       x = "Major Category",
       y = "Median ($)")
```

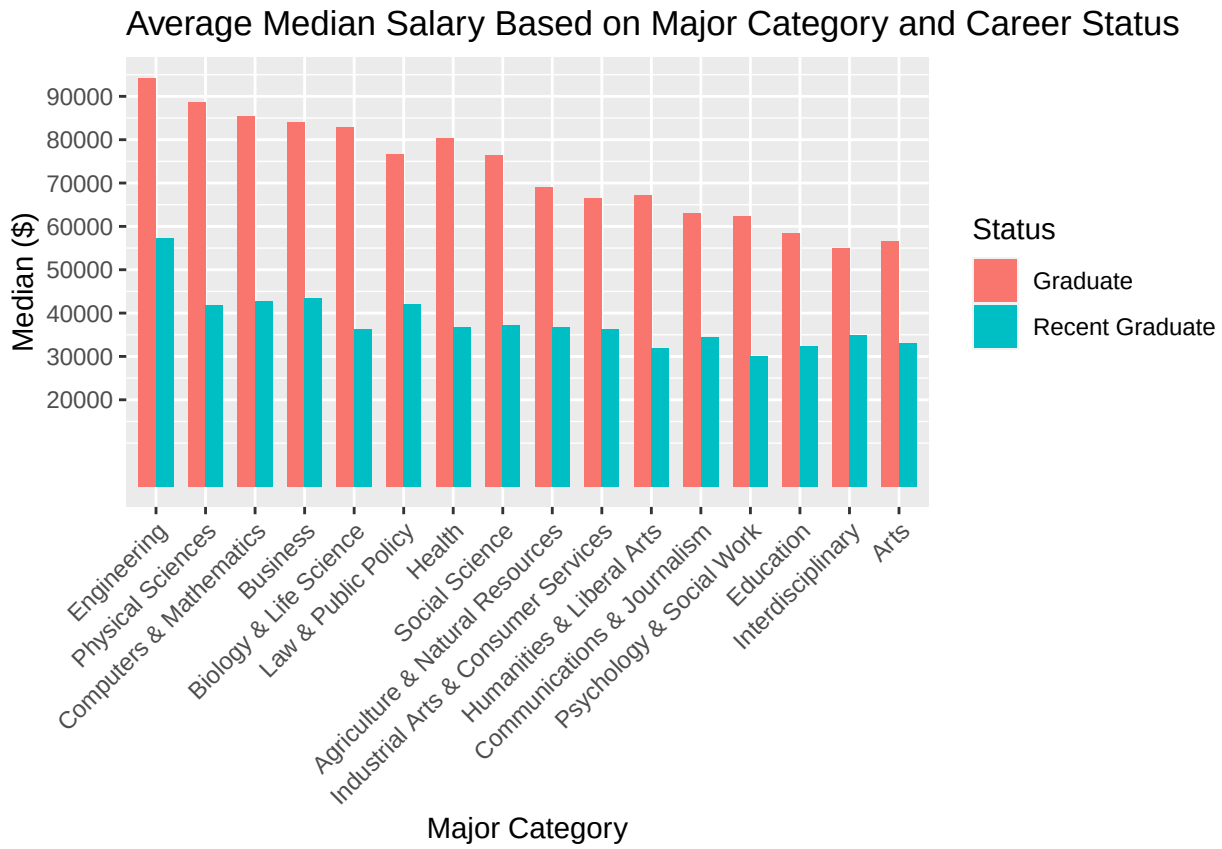
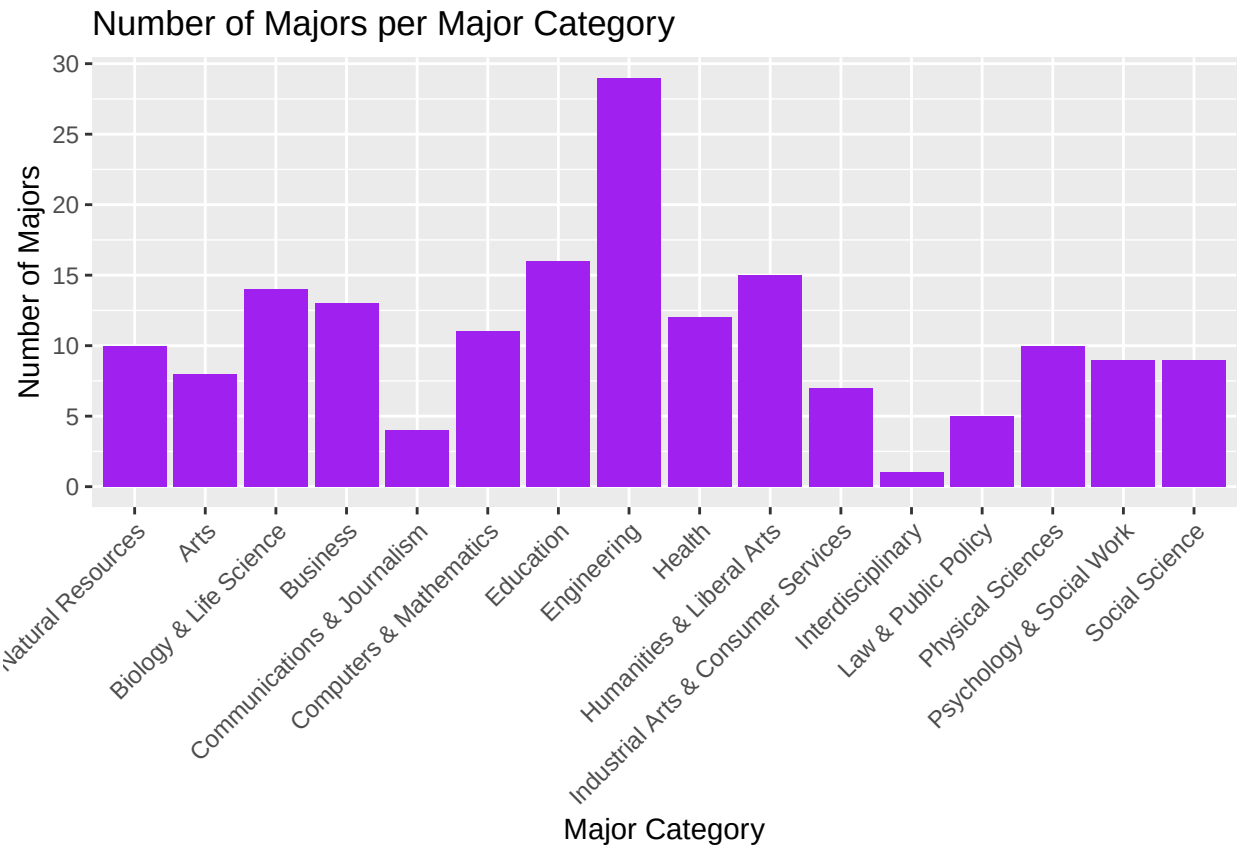


Figure 1: This plot depicts the average median salary of each major category with education status. We can see that on average, engineering majors have the highest starting salary as well as mid-career salary. In the next section, average median salary of starting salary and mid-career salary will be looked at independently to get a closer look at the salary distribution.

II) Distribution of Majors in Each Major Category

```r
# Plot of average median starting-career salary of each major category
n_grads %>%
  ggplot(aes(Major_category)) +
  geom_histogram(stat = "count", fill = "purple") +
```

```
  theme(axis.text.x = element_text(angle=45, hjust=1)) +
  scale_y_continuous(breaks = seq(0, 30, 5)) +
  labs(title = "Number of Majors per Major Category",
       x = "Major Category",
       y = "Number of Majors")
```




```
n_grads %>%
  group_by(Major_category) %>%
  summarize(count = n())
```

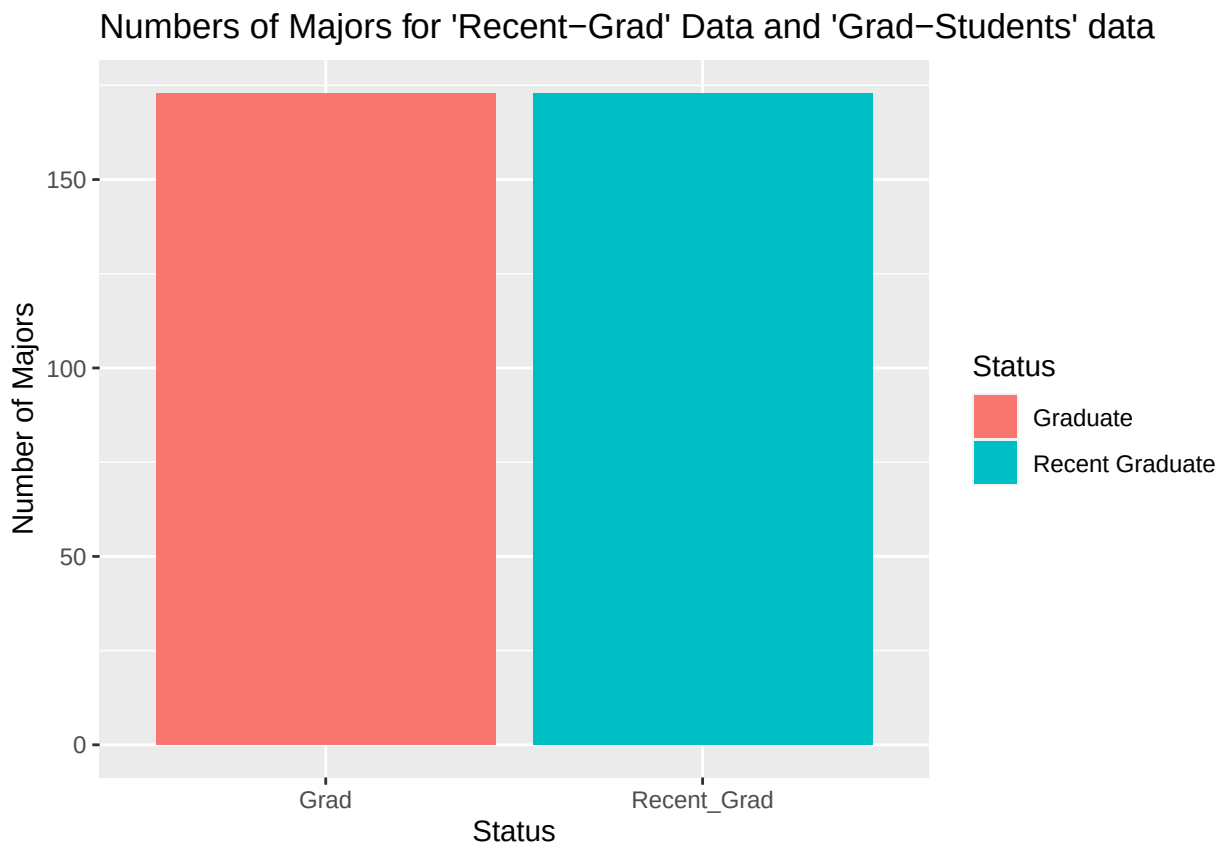
```
## # A tibble: 16 x 2
##    Major_category                     count
##    <chr>                              <int>
##  1 Agriculture & Natural Resources       10
##  2 Arts                                   8
##  3 Biology & Life Science                14
##  4 Business                              13
##  5 Communications & Journalism            4
##  6 Computers & Mathematics               11
##  7 Education                             16
##  8 Engineering                           29
##  9 Health                                12
## 10 Humanities & Liberal Arts             15
## 11 Industrial Arts & Consumer Services    7
```

```
## 12 Interdisciplinary                   1
## 13 Law & Public Policy                 5
## 14 Physical Sciences                  10
## 15 Psychology & Social Work            9
## 16 Social Science                      9
```

Figure 2: Histogram plot of the distribution of number of majors per major category. This plot shows that engineering has the most number of majors (29) while interdisciplinary has the least (1).

III) Distribution of 'Recent-Grads' and 'Grad-students'

```
# Plot of average median starting-career salary of each major category
new_combined_data %>%
  ggplot(aes(x = Status, fill = Status)) +
  geom_histogram(stat = "count", position = "dodge") +
  theme(axis.text.x = element_text(hjust=0.5)) +
  scale_fill_discrete(labels = c("Graduate", "Recent Graduate")) +
  labs(title = "Numbers of Majors for \'Recent-Grad\' Data and \'Grad-Students\' data",
       y = "Number of Majors",
       x = "Status")
```




```
new_combined_data %>%
  summarize(count_recent_grads = sum(Status == "Recent_Grad"),
            count_grads = sum(Status == "Grad"))
```

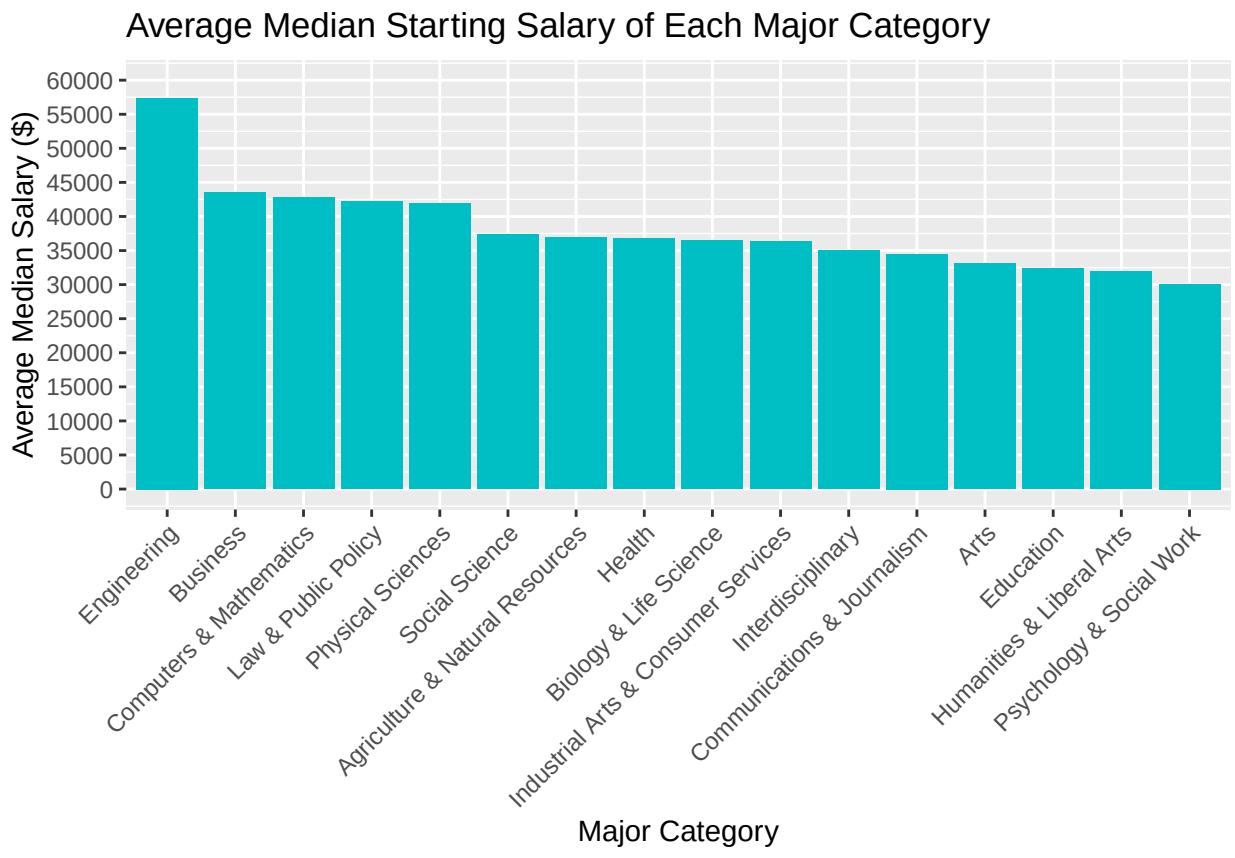
```
##   count_recent_grads count_grads
## 1                173         173
```

Figure 3: This plot shows the number of observations/majors for both 'recent-grad' data and 'grad-students' data. This depicts that there were the same number of majors from each dataset: 173 majors per status. This is important because when merging data and coming to conclusions with our data, we need to ensure that all the majors are the same and there are the same number of majors.

b. Salary Distributions

I) Starting Salary Distribution

```
# Plot of average median starting-career salary of each major category
new_combined_data %>%
  filter(Status == "Recent_Grad") %>%
  group_by(Major_category) %>%
  mutate(avg_median = mean(Median)) %>%
  ggplot(aes(x = reorder(Major_category, -avg_median), y = avg_median)) +
  geom_bar(stat = "summary", fun = "mean", fill = "#00BFC4") +
  scale_y_continuous(breaks = seq(0, 100000, 5000), limits = c(0, 60000)) +
  theme(axis.text.x = element_text(angle=45, hjust=1)) +
  labs(title = "Average Median Starting Salary of Each Major Category",
       y = "Average Median Salary ($)",
       x = "Major Category")
```



```r
# salary summary statistics
new_combined_data %>%
  filter(Status == "Recent_Grad") %>%
  group_by(Major_category) %>%
  summarize(avg_median = mean(Median),
            P25 = (avg_median * 2) * 0.25,
            P75 = (avg_median * 2) * 0.75) %>%
  arrange(desc(avg_median))
```

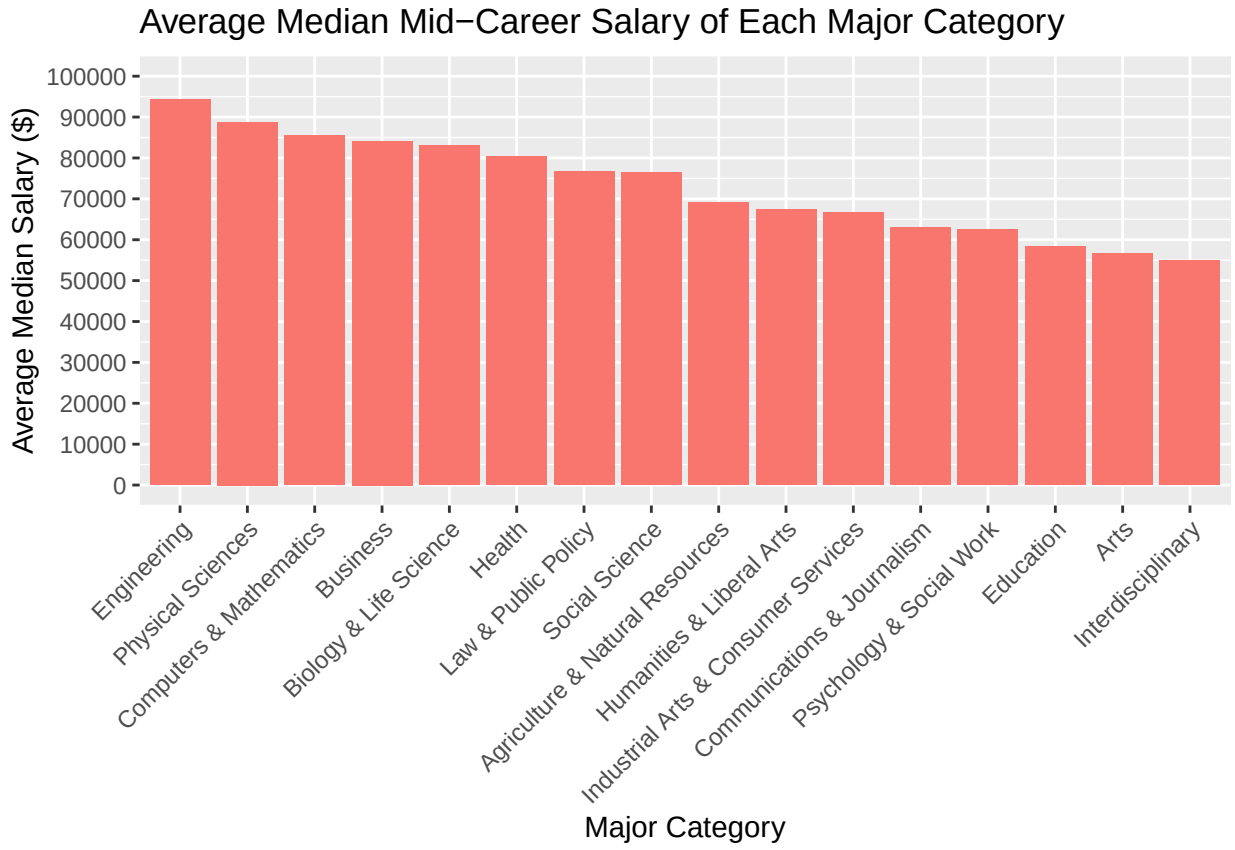
```
## # A tibble: 16 x 4
##    Major_category                      avg_median    P25    P75
##    <chr>                                    <dbl>  <dbl>  <dbl>
##  1 Engineering                             57383. 28691. 86074.
##  2 Business                                43538. 21769. 65308.
##  3 Computers & Mathematics                 42745. 21373. 64118.
##  4 Law & Public Policy                     42200  21100  63300
##  5 Physical Sciences                       41890  20945  62835
##  6 Social Science                          37344. 18672. 56017.
##  7 Agriculture & Natural Resources         36900  18450  55350
##  8 Health                                  36825  18412. 55238.
##  9 Biology & Life Science                  36421. 18211. 54632.
## 10 Industrial Arts & Consumer Services     36343. 18171. 54514.
## 11 Interdisciplinary                       35000  17500  52500
## 12 Communications & Journalism             34500  17250  51750
## 13 Arts                                    33062. 16531. 49594.
## 14 Education                               32350  16175  48525
## 15 Humanities & Liberal Arts               31913. 15957. 47870
## 16 Psychology & Social Work                30100  15050  45150
```

Figure 4: This plot depicts the average median starting salary of each major category. Engineering majors, on average, have a significantly higher starting median salary relative to other major categories. In fact, engineering majors have a starting median salary more than $13,000 above the next highest major category, which is business.

II) Mid-Career Salary Distribution

```r
# remove use of scientific notation
options(scipen = 999)

# Plot of average median mid-career salary of each major category
new_combined_data %>%
  filter(Status == "Grad") %>%
  group_by(Major_category) %>%
  mutate(avg_median = mean(Median)) %>%
  ggplot(aes(x = reorder(Major_category, -avg_median), y = avg_median)) +
  geom_bar(stat = "summary", fun = "mean", fill = "#F8766D") +
  scale_y_continuous(breaks = seq(0, 100000, 10000), limits = c(0, 100000))+
  theme(axis.text.x = element_text(angle=45, hjust=1)) +
  labs(title = "Average Median Mid-Career Salary of Each Major Category",
       y = "Average Median Salary ($)",
       x = "Major Category")
```



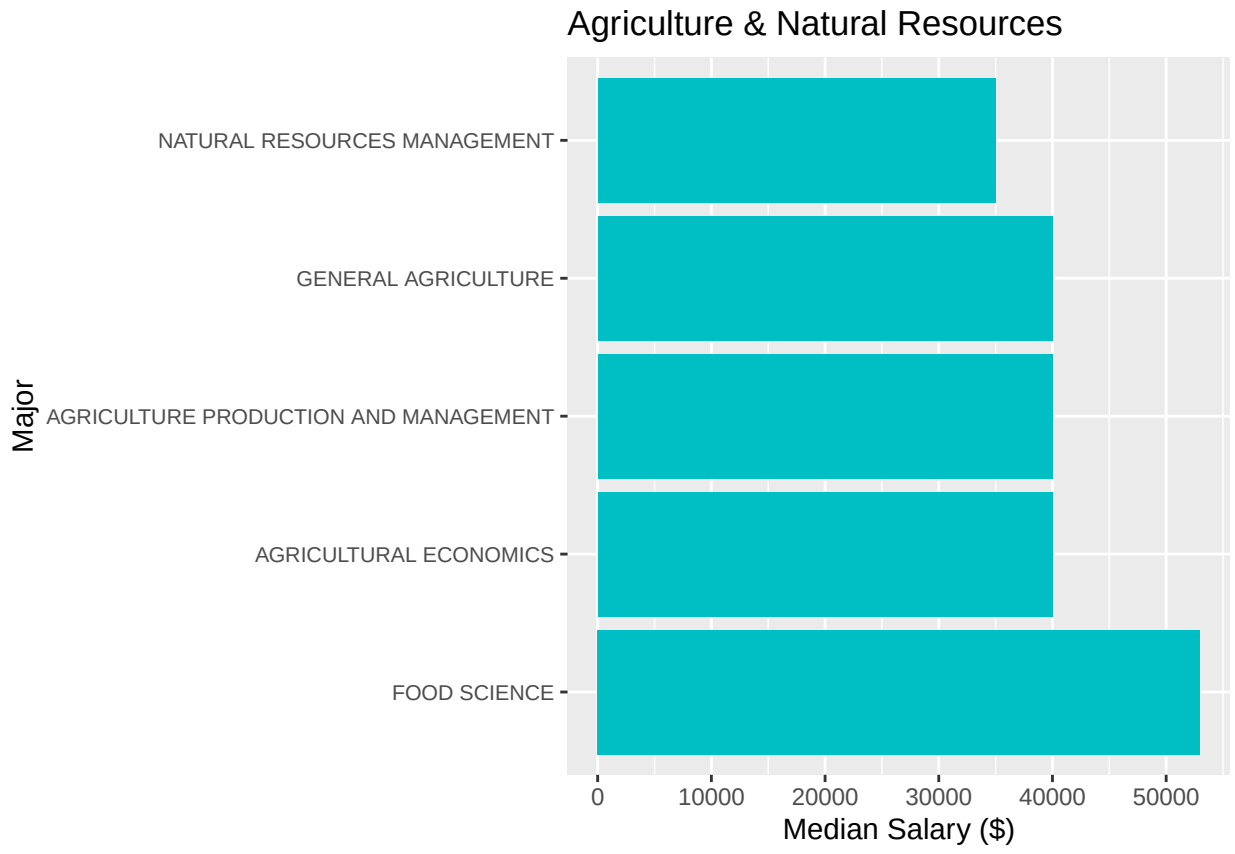

```
# average median salary of each major category
new_combined_data %>%
  filter(Status == "Grad") %>%
  group_by(Major_category) %>%
  summarize(avg_median = mean(Median)) %>%
  arrange(desc(avg_median))
```

```
## # A tibble: 16 x 2
##    Major_category                       avg_median
##    <chr>                                     <dbl>
##  1 Engineering                              94328.
##  2 Physical Sciences                        88800
##  3 Computers & Mathematics                  85545.
##  4 Business                                 84154.
##  5 Biology & Life Science                   83000
##  6 Health                                   80292.
##  7 Law & Public Policy                      76600
##  8 Social Science                           76444.
##  9 Agriculture & Natural Resources          69130
## 10 Humanities & Liberal Arts                67333.
## 11 Industrial Arts & Consumer Services      66571.
## 12 Communications & Journalism              63000
## 13 Psychology & Social Work                 62456.
## 14 Education                                58438.
## 15 Arts                                     56544.
## 16 Interdisciplinary                        55000
```

Figure 5: This plot depicts the average median mid-career salary for each major category. While engineering majors are still ranked highest, physical sciences majors are ranked 2 on this plot instead of business majors. In addition, the average median salaries between consecutive major categories are much closer. Whereas the average median salary difference between the rank 1 and rank 2 major categories was above $13,000 in the previous plot, it has reduced to less than $6,000 in this plot (between engineering and physical sciences).

III) Top 5 Majors of Each Major Category for Starting Salary

```
# Sample code for plot of top 5 majors by median starting salary in a major category
new_combined_data %>%
  filter(Status == "Recent_Grad", Major_category == "Agriculture & Natural Resources") %>%
  arrange(desc(Median)) %>%
  slice(1:5) %>%
  ggplot(aes(x = reorder(Major, -Median), y = Median)) +
  geom_bar(stat = "summary", fun = "mean", fill = "#00BFC4") +
  coord_flip() +
  theme(axis.text.y = element_text(size = 8)) +
  scale_y_continuous(breaks = seq(0, 50000, 10000)) +
  labs(title = "Agriculture & Natural Resources",
       x = "Major",
       y = "Median Salary ($)")
```



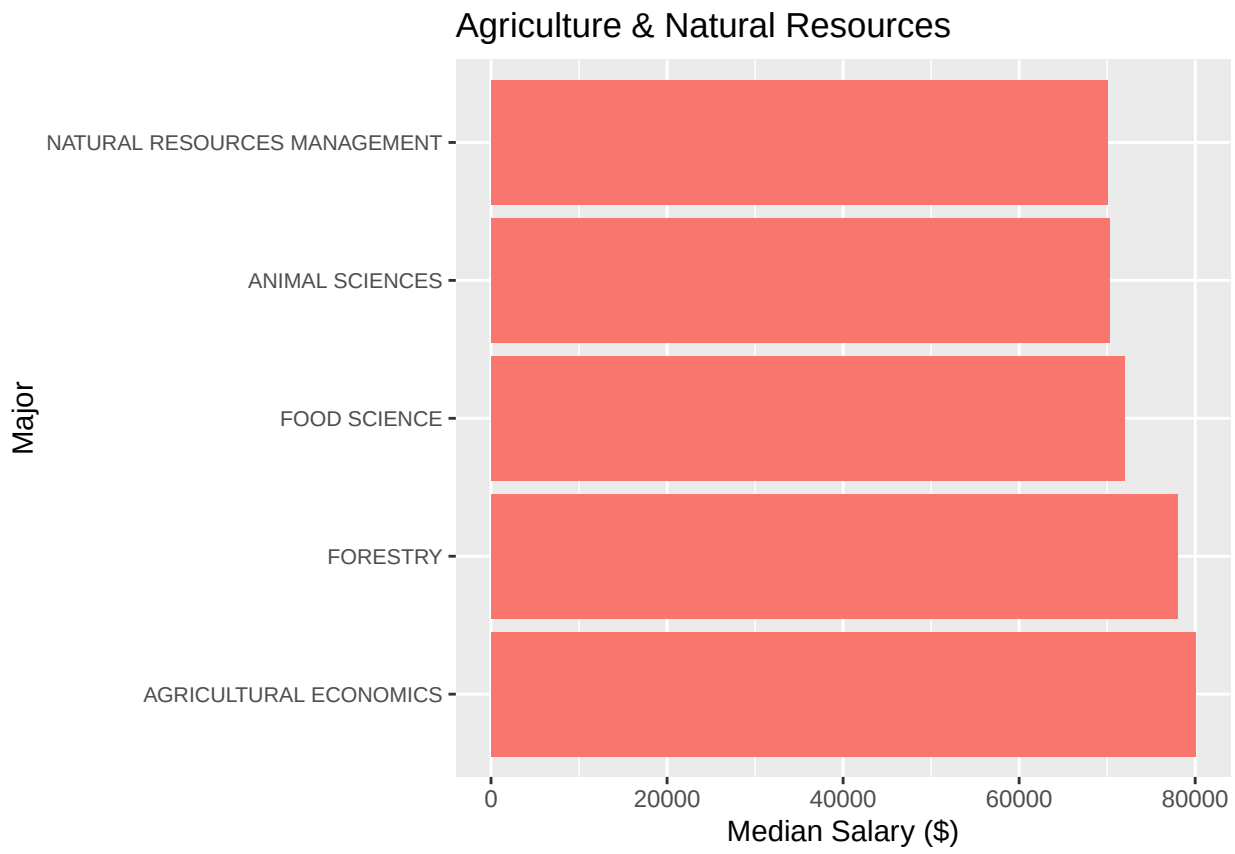
```
# code repeated for all major categories, only changing the major category name, and
# x-axis scale as needed
# plots uploaded to a doc and linked in supporting text below
```

Figure 6: After visualizing the distribution of average median starting salary for each major category, we wanted to see the majors with highest median starting salary for each of the major categories. Often times, choosing a degree and job isn't solely about the money. It is often a combination of what pays well in the field an individual is interested in. In order to answer this question, these plots depict the top 5 majors by median starting salary for each of the major categories.

The code provided is a sample code for just one of the major categories. Keeping the report length and flow in best interest, the plots for the other major categories have been aggregated into a google document (Link: Starting Salary Top 5 Majors) and can be viewed as intended.

IV) Top 5 Majors of Each Major Category for Mid-Career Salary

```
# Sample code for plot of top 5 majors by median starting salary in a major category
new_combined_data %>%
  filter(Status == "Grad", Major_category == "Agriculture & Natural Resources") %>%
  arrange(desc(Median)) %>%
  slice(1:5) %>%
  ggplot(aes(x = reorder(Major, -Median), y = Median)) +
  geom_bar(stat = "summary", fun = "mean", fill = "#F8766D") +
  coord_flip() +
  theme(axis.text.y = element_text(size = 8)) +
  scale_y_continuous(breaks = seq(0, 80000, 20000)) +
  labs(title = "Agriculture & Natural Resources",
       x = "Major",
       y = "Median Salary ($)")
```




```
# code repeated for all major categories, only changing the major category name, and
# x-axis scale as needed
# plots uploaded to a doc and linked in supporting text below
```

Figure 7: The plots for the top 5 majors in each major category for mid-career median salary was then visualized here.

As done previously, the code provided is a sample code for just one of the major categories. Again, keeping the report length and flow in best interest, the plots for the other major categories have been aggregated into a separate google document, (Link: Mid-Career Salary Top 5 Majors) and can be viewed as intended.

c. Comparison of Salary Growth

I) Salary Growth From Starting to Mid-Career for Each Major Category

```
# add percent change variable: change from starting to mid-career salary
combined_data <- combined_data %>%
  group_by(Major) %>%
  mutate(percent_change = ((Grad_median - Median) / Median) * 100) %>%
  mutate(across(percent_change, round, 2))

# add percent change variable by grouping major category
dfMC <- combined_data %>%
  group_by(Major_category.x) %>%
```

```
  mutate(mean_percent_change = mean(percent_change)) %>%
  mutate(across(mean_percent_change, round, 2))

# plot of percent change of each major category
ggplot(dfMC, aes(x = reorder(Major_category.x, mean_percent_change),
                 y = mean_percent_change)) +
  geom_bar(stat = "summary", fun = "mean", fill = "purple", alpha = 0.7) +
  geom_text(aes(label = mean_percent_change), size = 4, hjust = 1.1) +
  coord_flip() +
  labs(title = "Average Percent Growth",
       x = "Major Category",
       y = "Percent Growth (%)")
```

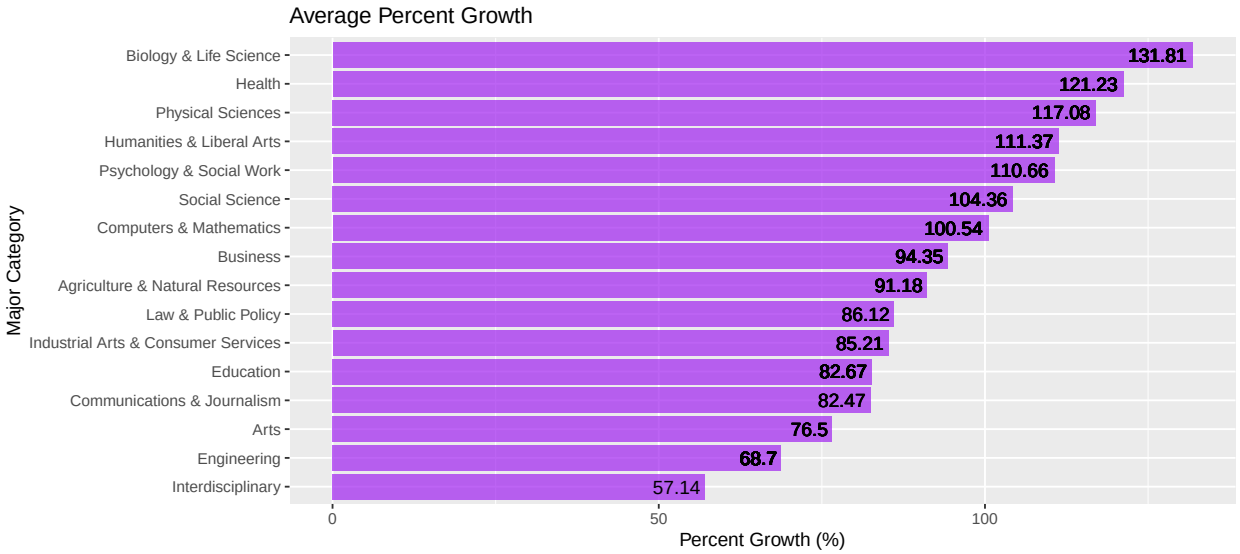


Figure 8: This first plot depicts the average percent growth of median salary from starting to mid-career for each of the major categories. This plot shows that biology and life science majors have, on average, the most percent growth from starting to mid-career salary at 131.81%. Though engineering majors had the highest average median starting and mid-career salaries, this plot shows that salary growth is low, at only 68.7%.

II) Top 20 Majors Ranked by Salary Growth

```
# arrange all majors by salary percent change and show top 20
dfOverall <- combined_data %>%
  arrange(desc(percent_change)) %>%
  head(20)

# plot of percent growth for the top 20 majors
options(scipen = 999)
ggplot(dfOverall, aes(x = reorder(Major, percent_change), Grad_median)) +
  geom_col(alpha = 0.5, fill = "red") +
  geom_col(aes(x = reorder(Major, percent_change), Median), alpha = 0.4, fill = "blue") +
  scale_y_continuous(breaks = seq(0, 140000, 20000)) +
```

```
geom_text(aes(label = percent_change), size = 5, hjust = 1.1) +
theme(axis.text.x = element_text(angle=0, size = 13),
      plot.title = element_text(size = 13),
      strip.text = element_blank(),
      axis.text.y.left = element_text(size = 11),
      axis.title.y = element_text(size = 18),
      axis.title.x = element_text(size = 18)) +
coord_flip() +
labs(title = "Top 20 Majors by Percent Growth from Starting to Mid-Career Salary",
     x = "Major",
     y = "Salary ($)")
```

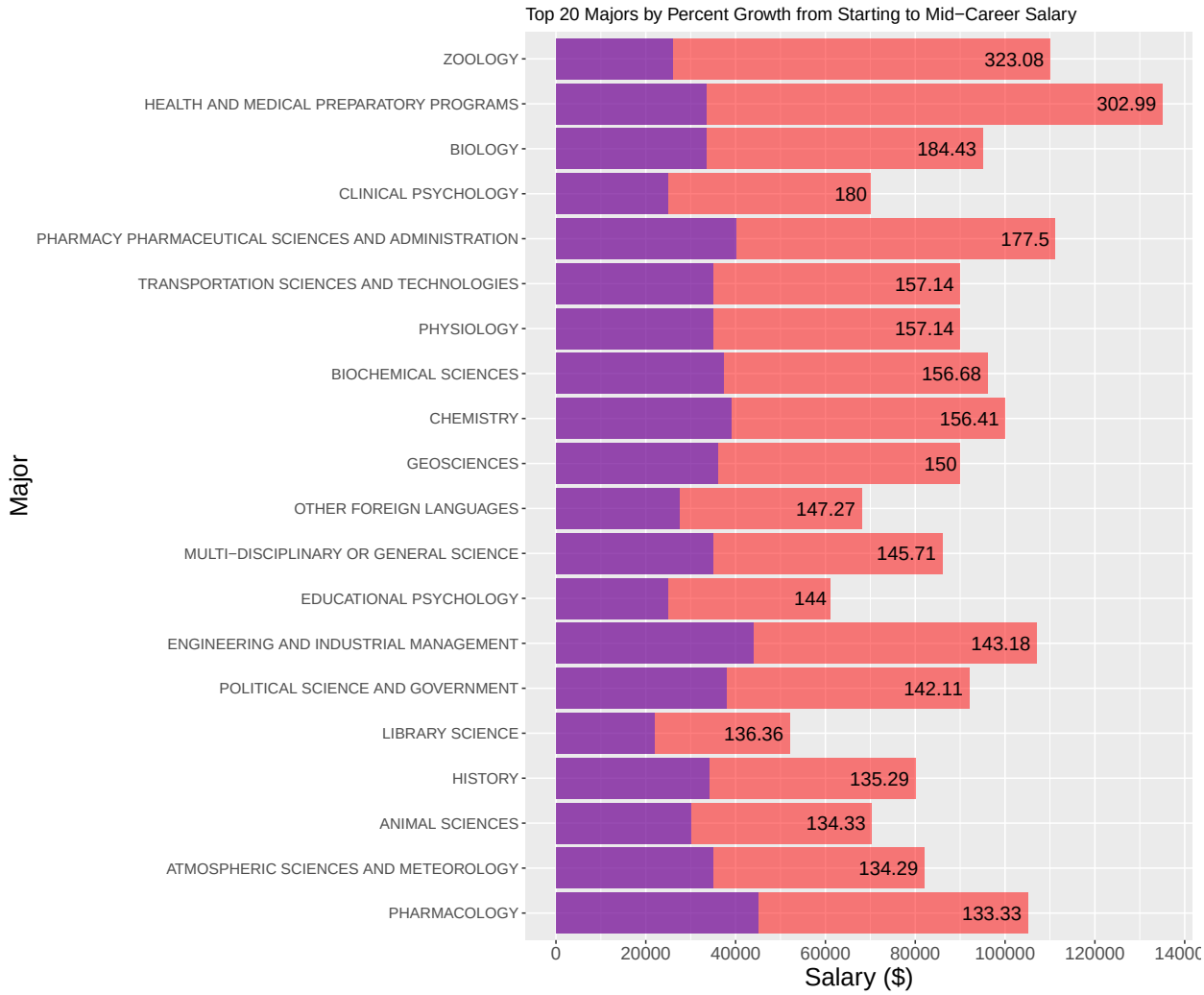


Figure 9: This second plot depicts the rankings of the top 20 majors by percent growth from starting to mid-career salary. The purple colored bars represent the starting salary for the given major, and the salmon colored bars represent the mid-career salary for the major. This plot shows that of all the majors in the dataset, zoology has the highest salary growth at 323.08%. This plot supports the findings of the previous plot, as most of the majors in the top 20 rankings are in the biology and life science major category, which had the highest average percent growth in plot 8.

III) Salary Growth of Each Major Category During 2009 - 2021

```
# change Major names into lowercase to match other datasets
# create a dataset with major categories and majors only
major_cat <- n_recent_grads %>%
  mutate(major = tolower(Major)) %>%
  mutate("Major" = str_to_title(major)) %>%
  select(Major_category, Major)

# joined the dataset of median salaries of majors from the year 2009 to 2021 with the dataset with majo
salary_all <- read.csv("~/Desktop/SDS322E/Project/project data/Salary 2009-2021.csv")
n_salary_all <- left_join(major_cat, salary_all, by = "Major")

# used pivot_longer to condense the columns into one column labeled year
t_salary_all <- pivot_longer(n_salary_all, cols = c('X2009', 'X2010', 'X2011', 'X2012',
                                                    'X2013', 'X2014', 'X2015', 'X2016',
                                                    'X2017', 'X2018', 'X2019', 'X2021'),
                             names_to = "year",
                             values_to = "Median Salary")
# remove coerced "X"s infront of each year
t_salary_all$year <- gsub("X", "", t_salary_all$year)


# Plot of average median salary over the years 2009 - 2021 for each major category
t_salary_all %>%
  group_by(Major_category, year) %>%
  na.omit() %>%
  mutate(avg_salary = mean(`Median Salary`)) %>%
  ggplot(aes(x = as.numeric(year), y = `avg_salary`, color = Major_category)) +
  geom_point(stat = "summary", fun = "mean", size = 2) +
  geom_line() +
  scale_x_continuous(breaks = seq(2009, 2021, 1), limits = c(2009, 2021)) +
  scale_y_continuous(breaks = seq(30000, 110000, 5000)) +
  theme(axis.text.x = element_text(size = 11),
        plot.title = element_text(size = 15),
        legend.text = element_text(size = 12),
        axis.text.y.left = element_text(size = 11),
        axis.title.y = element_text(size = 13),
        axis.title.x = element_text(size = 13)) +
  labs(title = "Average Median Salary Over 2009 - 2021 for Each Major Category",
       x = "Year",
       y = "Average Median Salary ($)")
```

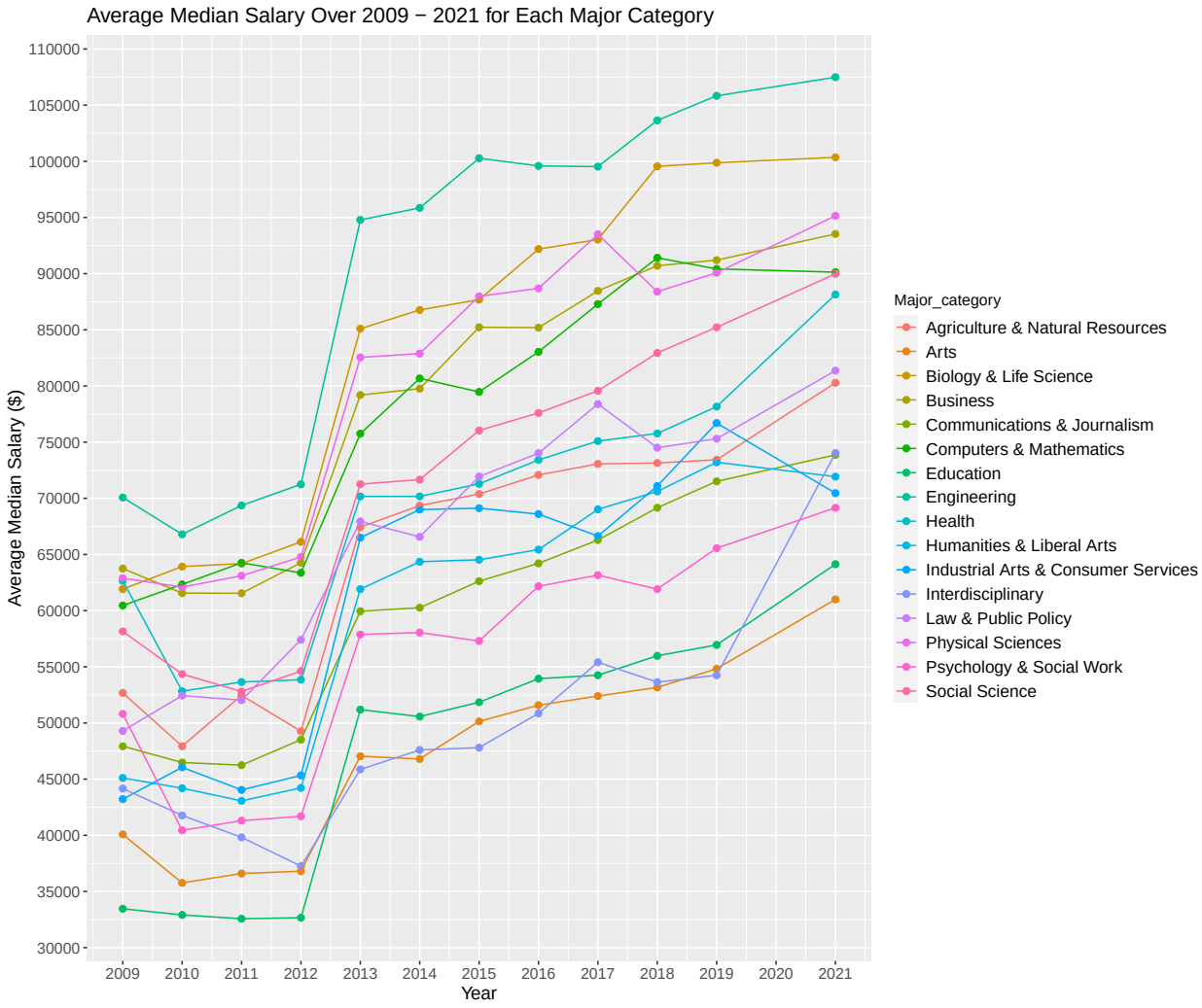


Figure 10: This plot depicts how the salary for each major category has changed from year to year over the duration of 2009 - 2021. The notable change observed in this plot is the spike in average median salary from 2012 - 2013. This time period lines up with recovery from the US housing market crash. Overall, the percent change in salary seems to be similar over time for each of the major categories. Therefore, this data suggests that economic conditions is one of the most impactful factors in median salary changes.

d. Unemployment

I) Unemployment Rate of Each Major Category

```
# merge the datasets by Major
unem_rates_all <- read.csv("~/Desktop/SDS322E/Project/project data/Unemployment Rates 2009-2021.csv")
n_unem_rates <- left_join(major_cat, unem_rates_all, by = "Major" )

# condense the columns into one column labeled year
t_unem_rates <- pivot_longer(n_unem_rates, cols = c('X2009', 'X2010', 'X2011', 'X2012',
                                                    'X2013', 'X2014', 'X2015', 'X2016',
                                                    'X2017', 'X2018', 'X2019', 'X2021'),
```

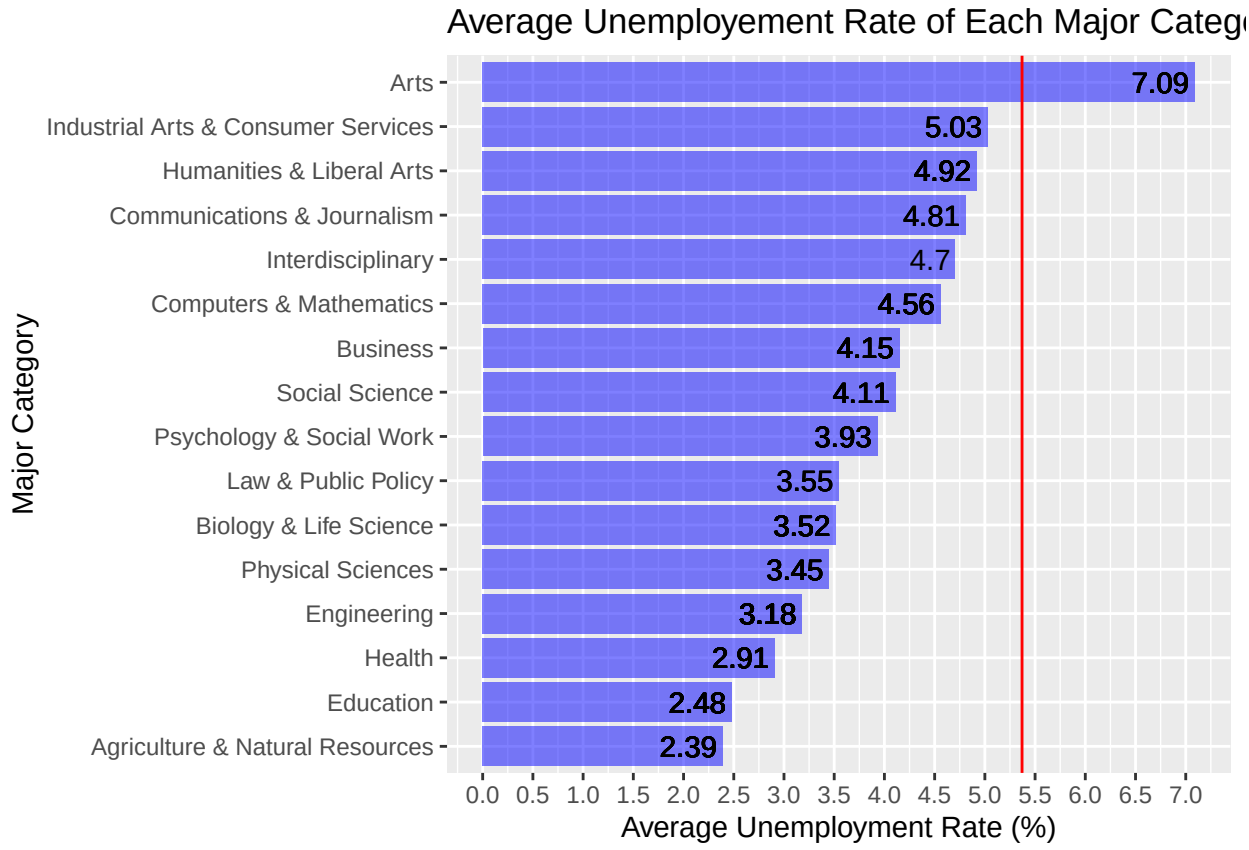
```
                                          names_to = "year",
                                          values_to = "Unemployment Rates")
# remove coerced "X"s infront of each year
t_unem_rates$year <- gsub("X", "", t_unem_rates$year)

# filter for year 2021 and save dataset
c_unem_rates <- t_unem_rates %>%
  filter(year == 2021) %>%
  na.omit() %>%
  group_by(Major_category) %>%
  mutate(avg_unem_rate = mean(`Unemployment Rates`)) %>%
  mutate(across(avg_unem_rate, round, 2))

# download US unemployment data
USunem <- read.csv("~/Desktop/SDS322E/Project/project data/US_unemployment.csv")
# pivot table to make it tidy and rename
USunem <- USunem %>%
  pivot_longer(cols = c("Jan":"Dec")) %>%
  rename(Month = "name",
         Unemployment_rate = "value")
# calculate average US unemployment rate
USunem %>%
  filter(Year == 2021) %>%
  summarize(avg = mean(Unemployment_rate))
```

```
## # A tibble: 1 x 1
##     avg
##   <dbl>
## 1  5.37
```

```
# plot of average unemployment rate in 2021 for each major category with verticle line
# represent overall US unemployment rate in 2021
c_unem_rates %>%
  ggplot(aes(x = reorder(Major_category, avg_unem_rate),
             y = avg_unem_rate)) +
  geom_bar(stat = "summary", fun = "mean", fill = "blue", alpha = 0.5) +
  geom_hline(yintercept=5.37,color="red") +
  geom_text(aes(label = avg_unem_rate), hjust = 1.1) +
  scale_y_continuous(breaks = seq(0, 8, 0.5)) +
  coord_flip() +
  labs(title = "Average Unemployement Rate of Each Major Category",
       x = "Major Category",
       y = "Average Unemployment Rate (%)")
```




```
# summary statistics
# categorize average unemployment as high if above national average, low if below
c_unem_rates %>%
  group_by(Major_category) %>%
  summarize(avg_unemployment_rate = mean(`Unemployment Rates`),
            ifelse(avg_unemployment_rate > 5.37, "High", "Low")) %>%
  arrange(desc(avg_unemployment_rate)) %>%
  rename(Unemployment_Severity = "ifelse(avg_unemployment_rate > 5.37, \"High\", \"Low\")")
```

```
## # A tibble: 16 x 3
##    Major_category                       avg_unemployment_rate Unemployment_Seve~1
##    <chr>                                                <dbl> <chr>
##  1 Arts                                                  7.09 High
##  2 Industrial Arts & Consumer Services                   5.03 Low
##  3 Humanities & Liberal Arts                             4.92 Low
##  4 Communications & Journalism                           4.81 Low
##  5 Interdisciplinary                                     4.70 Low
##  6 Computers & Mathematics                               4.56 Low
##  7 Business                                              4.15 Low
##  8 Social Science                                        4.11 Low
##  9 Psychology & Social Work                              3.93 Low
## 10 Law & Public Policy                                   3.55 Low
## 11 Biology & Life Science                                3.52 Low
## 12 Physical Sciences                                     3.45 Low
## 13 Engineering                                           3.18 Low
```

```
## 14 Health                                                2.91 Low
## 15 Education                                             2.48 Low
## 16 Agriculture & Natural Resources                       2.39 Low
## # ... with abbreviated variable name 1: Unemployment_Severity
```

Figure 11: This plot depicts the average unemployment of each major category in 2021. The red vertical line represents the overall unemployment rate in the US in 2021. This plot indicated that arts majors were the only major category to have an unemployment rate above the national average. This is likely because careers in art may not have typical employers. On the other hand, all the other major categories had unemployment rates below the national average. This is likely due to the fact that having some sort of degree places an individual at an advantage when obtaining a job compared to an individual without a college degree.

3. Discussion

a. Research Question 1: Which college degree should you pursue for a career with a high starting and mid-career salary?

The data indicated that pursuing a degree in an engineering major is most beneficial to obtain a high starting as well as mid-career salary, as depicted in figures 4 and 5. The average median starting salary for the engineering major category was $57, 382, and the average median mid-career salary was $94,327.59 (2010 - 2012). However, for individuals who have interests in other fields, it would be more helpful to look at the top paying majors for each major category. For example, as depicted in figure 6, those who have an interest in the agriculture & natural resources field should consider a degree in food science to prioritize a high starting salary. On the other hand, as depicted in figure 7, those prioritizing a high mid-career salary (in the same field) should consider a degree in agricultural economics.

b. Research Question 2: Which college degree shows the most promising growth rate from starting to mid-career salary as well as in respect to time?

A degree in biology and life sciences was shown to be have the highest growth rate from starting to mid-career median salary at 131.81% (figure 8). When compared with all majors in our dataset, zoology was shown to have the highest percent growth at 323.08% (figure 9), which unsurprisingly is in the biology and life sciences major category.

In terms of salary growth in respect to time, our data showed that salary growth was similar for the major categories during the years 2009 - 2021 (figure 10). Much of the pattern of salary change for each major category as shown in figure 10 is similar, including the notable spike in median salary that all major categories experienced in 2012. Therefore, as noted prior, this suggests that overall economic trends are a strong predictor for median salary growth from year to year for most majors.

c. Reflections

The most valuable skill we learned from this project is the process of defining a research question(s) to investigate, finding the data necessary to address the problem, and designing the work flow to execute the report. This process was time consuming and difficult at times, but finding ways to overcome these challenges gave us valuable skill sets we can apply in the future. The most challenging aspect of this project was defining the research question as well as finding the data itself. In addition, methods of coding we have not covered in class often had to be utilized in this project, which was challenging, but was also a reason that made this project fun.

d. Acknowledgements

This project was completed by the collaborative efforts of Jongho Yoo and Hyunji Kang. All parts of the project, including research, data collection, code, and writing were distributed as fairly and evenly as possible.