3_10_2021_DSA5200_Final Project

Owen R. Evans

3/10/2021

Data Description

The following visualization project utilizes data obtained from the American Community Survey 2010-2012 Public Use Microdata Series. This dataset provides information on the starting salaries, unemployment rate and underemployment rates for recent college graduates from 2010-2012. The dataset includes interesting information regarding the influence of college major choice on earnings potential, broken down by gender participation. It is important to note that the starting salaries reported were highly influenced by the economic climate at the time.

Unemployment rates per college major and gender were acquired via random sampling of larger groups and thus may be subject to error. Underemployment rates per major category were calculated as the quotient of those employed in jobs requiring a degree over the sum of those in low-wage jobs and jobs not requiring a degree. The dataset had one NA that was removed for the visualization and no obvious outliers.

```
# Load College Major Data
women.stem <- read.csv(paste0("https://raw.githubusercontent.com/fivethirtyeight",
    "/data/master/college-majors/women-stem.csv"))  # women in STEM degrees

recent.grads <- read.csv(paste0("https://raw.githubusercontent.com/fivethirtyeight",
    "/data/master/college-majors/recent-grads.csv"))  # All data for recent graduates

grad.students <- read.csv(paste0("https://raw.githubusercontent.com/fivethirtyeight",
    "/data/master/college-majors/grad-students.csv"))  # graduate students
```

Choosing a College Major

What are the specific majors with the most and least highest starting salaries?

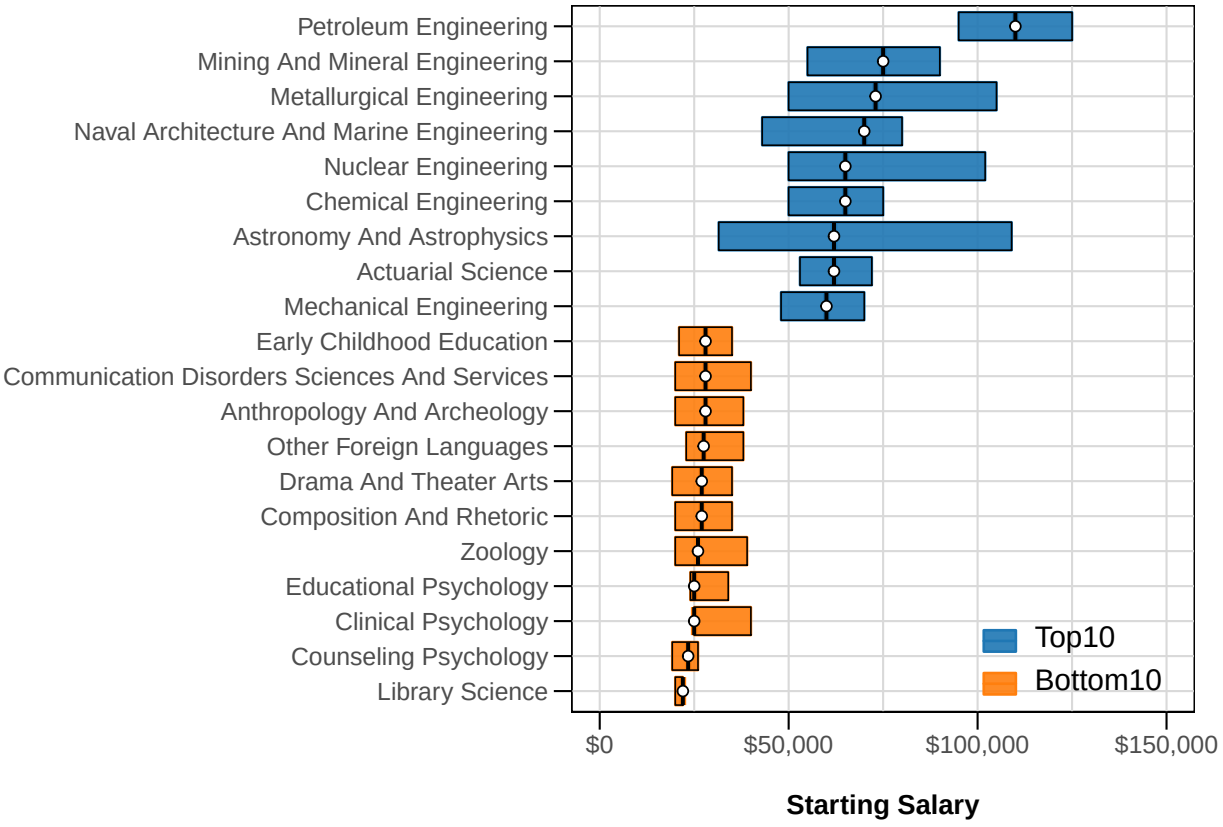
Figure 1

```
# Top 10 and Bottom 10 majors, earning power

sliced_data <- recent.grads %>% filter(Rank < 10 |
    Rank %in% c((nrow(recent.grads) - 10):(nrow(recent.grads)))) %>%
    mutate(Group = ifelse(Rank < 10, "Top10", "Bottom10")) %>%
    select(Rank, Group, Major, Major_category, Median,
        P25th, P75th)
```

```r
sliced_data$Major <- str_to_title(sliced_data$Major)
sliced_data$Group <- factor(sliced_data$Group, c("Top10",
    "Bottom10"))

ggplot(sliced_data, aes(x = reorder(Major, Median),
    y = Median, ymin = P25th, ymax = P75th)) + geom_crossbar(aes(color = Group,
    fill = Group), size = 0.5, width = 0.8, alpha = 0.9) +
    geom_crossbar(color = "black", size = 0.3, width = 0.8) +
    geom_point(color = "black", size = 1.7, shape = 21,
        fill = "white") + labs(x = NULL) + coord_flip() +
    scale_y_continuous(labels = dollar_format(), limits = c(0,
        150000)) + scale_fill_d3() + scale_color_d3() +
    theme_minimal() + theme(axis.title = element_text(size = 11,
    face = "bold"), plot.title = element_text(size = 13,
    face = "bold", hjust = 1), plot.subtitle = element_text(size = 11,
    hjust = 1, face = "italic"), plot.caption = element_text(size = 8,
    color = "grey65", hjust = 0.9), axis.text.y = element_text(size = 10),
    axis.text.x = element_text(size = 10), axis.line = element_blank(),
    panel.border = element_rect(color = "black", fill = NA,
        size = 0.35), axis.ticks = element_line(colour = "black",
        size = 0.4), axis.ticks.length = unit(0.25,
        "cm"), panel.grid.major = element_line(size = 0.4,
        color = "gray85"), panel.grid.minor = element_line(size = 0.4,
        color = "gray85"), legend.position = c(0.8,
        0.08), legend.title = element_blank(), legend.text = element_text(size = 12,
        vjust = 0.7), legend.background = element_blank(),
    panel.background = element_rect(fill = "white"),
    plot.margin = unit(c(0, 0.8, 0, 0), "cm")) + labs(subtitle = "What are the Most and Least Lucrative
    y = "\nStarting Salary", caption = "\nData Source : https://github.com/fivethirtyeight/data/blob/ma
    title = "Median and Interquartile Ranges of Starting Salaries for Recent Graduates")
```

Median and Interquartile Ranges of Starting Salaries for Recent Graduates

What are the Most and Least Lucrative College Majors?



Data Source : https://github.com/fivethirtyeight/data/blob/master/college-majors/recent-grads.csv

Choosing a College Major - Are Popular Majors a Good Route?

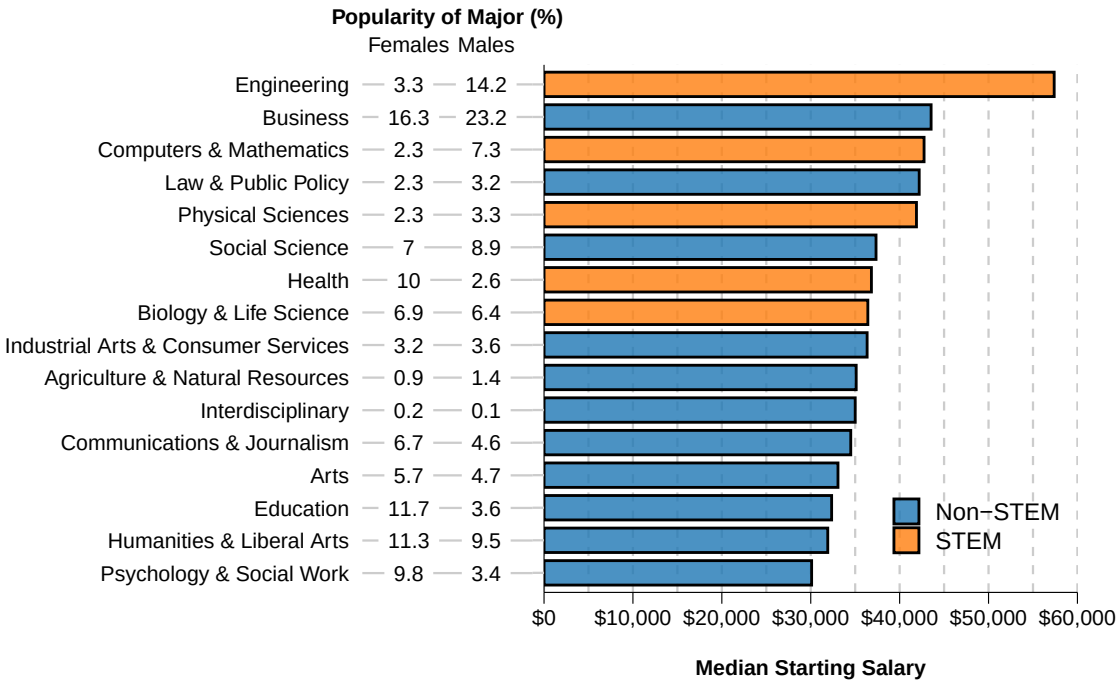
Should one choose a major based on popularity and are popular majors lucrative?

Figure 2

```
# Data preparation, remove NA's Added popularity
# index by Gender for each major
recent.grads <- na.omit(recent.grads)  # remove NA's
recent.grads$Total <- as.numeric(recent.grads$Total)
popular.data <- recent.grads %>% mutate(popIndexFemale = recent.grads$Women/sum(recent.grads$Women) *
    100, popIndexMale = recent.grads$Men/sum(recent.grads$Men) *
    100)


# Grouped by Major Category, Popularity % Split by
# Gender Aggregated data
popdata <- recent.grads %>% group_by(Major_category) %>%
    summarize(medsal = mean(Median), Total2 = sum(Total),
```

```
        Men2 = sum(Men), Women2 = sum(Women)) %>% mutate(popIndexMale = Men2/sum(Men2) *
    100, popIndexFemale = Women2/sum(Women2) * 100,
    Cat2 = ifelse(Major_category %in% women.stem$Major_category,
        "STEM", "Non-STEM")) %>% arrange(medsal)

# Reorder/re-factor major categor by decreasing
# median salary
popdata$Major_category <- factor(popdata$Major_category,
    levels = popdata$Major_category)

# Subplots - bar chart of median salary, ranked.
Mainbar <- ggplot(popdata, aes(x = Major_category,
    y = medsal)) + geom_bar(stat = "identity", aes(fill = Cat2),
    color = "black", width = 0.75, alpha = 0.8) + scale_y_continuous(labels = dollar_format(),
    limits = c(0, 60000), expand = c(0, 0)) + scale_fill_d3() +
    coord_flip() + theme_minimal() + theme(axis.title.y = element_blank(),
    plot.caption = element_text(size = 5, color = "grey33",
        hjust = 1), axis.text.y = element_blank(),
    axis.text.x = element_text(size = 8, color = "black"),
    axis.title.x = element_text(size = 8, face = "bold"),
    legend.title = element_blank(), axis.ticks.x = element_line(color = "black",
        size = 0.2), axis.ticks.y = element_line(color = "grey80",
        size = 0.4), axis.ticks.length = unit(0.15,
        "cm"), legend.text = element_text(size = 9,
        color = "black"), legend.key.size = unit(0.75,
        "lines"), axis.line = element_line(linetype = "solid",
        size = 0.2), panel.grid.major.x = element_line(size = 0.4,
        color = "gray80", linetype = "dashed"), panel.grid.minor.x = element_line(size = 0.4,
        color = "gray80", linetype = "dashed"), panel.grid.major.y = element_blank(),
    plot.margin = unit(c(1.11, 1, 0, 0), "cm"), legend.position = c(0.81,
        0.14)) + labs(y = "\nMedian Starting Salary\n",
    caption = paste0("\nData Source :", "https://github.com/fivethirtyeight/",
        "data/blob/master/college-majors/recent-grads.csvNote: STEM-Science, Technology, Engineering & M

# Pivot longer in order to derive datatable
popdata2 <- popdata %>% pivot_longer(cols = starts_with("popIndex"),
    names_to = "Sex", values_to = "PopularIndex")
# Rename variables
popdata2$Sex <- ifelse(popdata2$Sex == "popIndexMale",
    "Males", "Females")

# GGplot - 'datatable'
datatable <- ggplot(popdata2, aes(x = Major_category,
    y = factor(Sex), label = format(value, 1), nsmall = 1)) +
    geom_label(label = round(popdata2$PopularIndex,
        digits = 1), size = 2.8, color = "black", label.size = NA) +
    scale_y_discrete(position = "right") + theme_minimal() +
    theme(axis.title.y = element_blank(), axis.text.y = element_text(size = 8,
        color = "black"), axis.text.x = element_text(size = 8,
        color = "black", hjust = 0.5), plot.margin = unit(c(0.005,
        0, 2.009, 0.05), "cm"), axis.title = element_text(size = 8,
        color = "black", face = "bold", hjust = 0.5),
        panel.background = element_rect(fill = "white",
```

```
            color = "white"), panel.grid.major.x = element_blank(),
        panel.grid.major.y = element_line(color = "grey80",
            size = 0.4)) + labs(y = "\nPopularity of Major (%)") +
    coord_flip()

combotitle <- "    College Major Popularity and Starting Salary"
combosubtitle <- "      Are Popular Majors the Most Lucrative?"

tgrobtitle <- text_grob(combotitle, size = 12, face = "bold",
    hjust = 0.015)
tgrobsubtitle <- text_grob(combosubtitle, size = 10,
    hjust = -0.1, face = "italic")

title <- as_ggplot(tgrobtitle) + theme(plot.margin = margin(0,
    1.8, 0, 0, "cm"))
subtitle <- as_ggplot(tgrobsubtitle) + theme(plot.margin = margin(0,
    0, 0, 0, "cm"))

ggarrange(title, NULL, subtitle, NULL, datatable, Mainbar,
    labels = NULL, widths = c(2.4, 2.8), heights = c(0.25,
        0.25, 3.8), ncol = 2, nrow = 3)
```



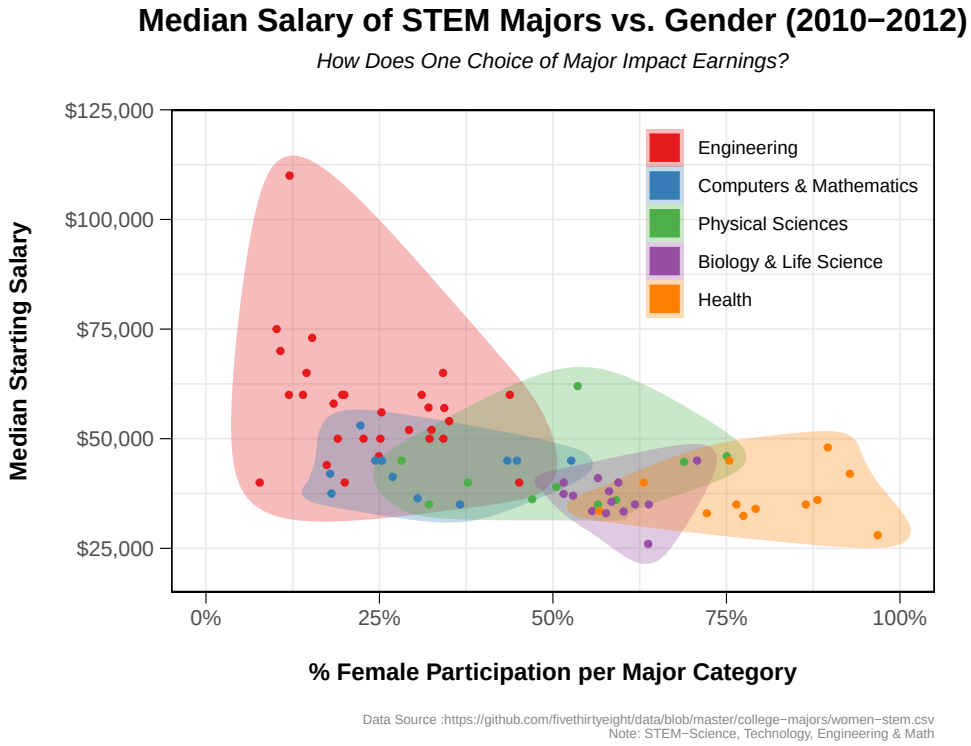
Data Source :https://github.com/fivethirtyeight/data/blob/master/college−majors/recent−grads.csvNote: STEM−Science, Technology, Engineering & Math

Should one choose a field in Science, Technology, Engineering and Math?

What is right college major choice for females?

Figure 3

```
# Data preparation
ranked <- women.stem %>% group_by(Major_category) %>%
    summarize(ShareWomen = mean(ShareWomen)) %>% arrange(ShareWomen)

women.stem$Major_category <- factor(women.stem$Major_category,
    levels = ranked$Major_category)

# Plot generation
ggplot(women.stem, aes(x = ShareWomen, y = Median,
    color = Major_category)) + geom_encircle(expand = 0.03,
    size = 5, aes(fill = Major_category), alpha = 0.3) +
    scale_color_brewer(palette = "Set1") + scale_fill_brewer(palette = "Set1") +
    geom_point(size = 0.8) + theme_minimal() + scale_x_continuous(labels = label_percent(),
    limits = c(0, 1)) + scale_y_continuous(labels = dollar_format(),
    limits = c(20000, 120000)) + theme(axis.title = element_text(size = 9,
    face = "bold"), plot.title = element_text(size = 12,
    face = "bold", hjust = 0.5), plot.subtitle = element_text(size = 8,
    hjust = 0.5, face = "italic"), plot.caption = element_text(size = 5,
    color = "gray53"), axis.text = element_text(size = 8),
    axis.ticks = element_line(colour = "black", size = 0.2),
    axis.ticks.length = unit(0.15, "cm"), panel.grid.major = element_line(size = 0.3),
    panel.grid.minor = element_line(size = 0.3), legend.position = c(0.8,
        0.78), legend.title = element_blank(), legend.text = element_text(size = 7),
    panel.border = element_rect(color = "black", fill = NA,
        size = 0.5), plot.margin = unit(c(0, 1.5, 0,
        1.5), "cm"), legend.key.size = unit(0.02, "lines")) +
    labs(title = "\nMedian Salary of STEM Majors vs. Gender (2010-2012)",
        subtitle = "How Does One Choice of Major Impact Earnings?\n",
        x = "\n% Female Participation per Major Category",
        y = "Median Starting Salary\n", caption = paste0("\nData Source :",
            "https://github.com/fivethirtyeight/data/blob",
            "/master/college-majors/women-stem.csv",
            "\nNote: STEM-Science, Technology, Engineering & Math"))
```



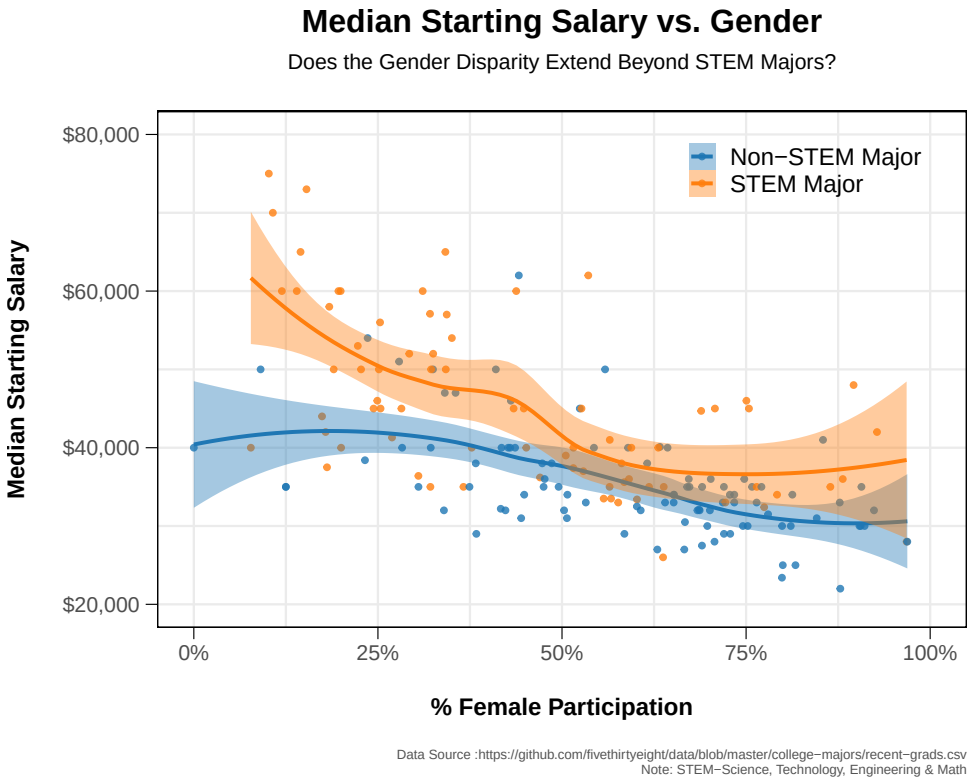
Gender vs. Starting Salary

Does a Gender Bias in starting salary exist for all college major categories?

Figure 4

```
recent.grads1 <- recent.grads %>% mutate(Cat2 = ifelse(Major_category %in%
    women.stem$Major_category, "STEM Major", "Non-STEM Major"))

# Percent Female vs Median Salary for All Majors
ggplot(recent.grads1, aes(x = ShareWomen, y = Median,
    color = Cat2)) + geom_point(size = 0.7, alpha = 0.8) +
    geom_smooth(aes(fill = Cat2, color = Cat2), size = 0.7) +
    theme_minimal() + scale_x_continuous(labels = label_percent(),
    limits = c(0, 1)) + scale_y_continuous(labels = dollar_format(),
    limits = c(20000, 80000)) + scale_fill_d3() + scale_color_d3() +
    theme(axis.title = element_text(size = 9, face = "bold"),
        plot.title = element_text(size = 12, face = "bold",
            hjust = 0.5), plot.subtitle = element_text(size = 8,
            hjust = 0.5), plot.caption = element_text(size = 5,
            color = "grey33"), axis.text = element_text(size = 8),
        axis.ticks = element_line(color = "black",
            size = 0.2), axis.ticks.length = unit(0.15,
            "cm"), panel.grid.major = element_line(size = 0.4),
        panel.grid.minor = element_line(size = 0.4),
        legend.position = c(0.8, 0.9), legend.title = element_blank(),
```

```
        legend.text = element_text(size = 9), plot.margin = unit(c(0,
            0, 0, 0), "cm"), panel.border = element_rect(color = "black",
            fill = NA, size = 0.5), legend.key.size = unit(0.7,
            "lines")) + labs(title = "Median Starting Salary vs. Gender",
    subtitle = "Does the Gender Disparity Extend Beyond STEM Majors?\n",
    x = "\n% Female Participation", y = "Median Starting Salary\n",
    caption = paste0("\nData Source :", "https://github.com/fivethirtyeight/",
        "data/blob/master/college-majors/recent-grads.csv
       Note: STEM-Science, Technology, Engineering & Math"))
```



Under- and Unemployment

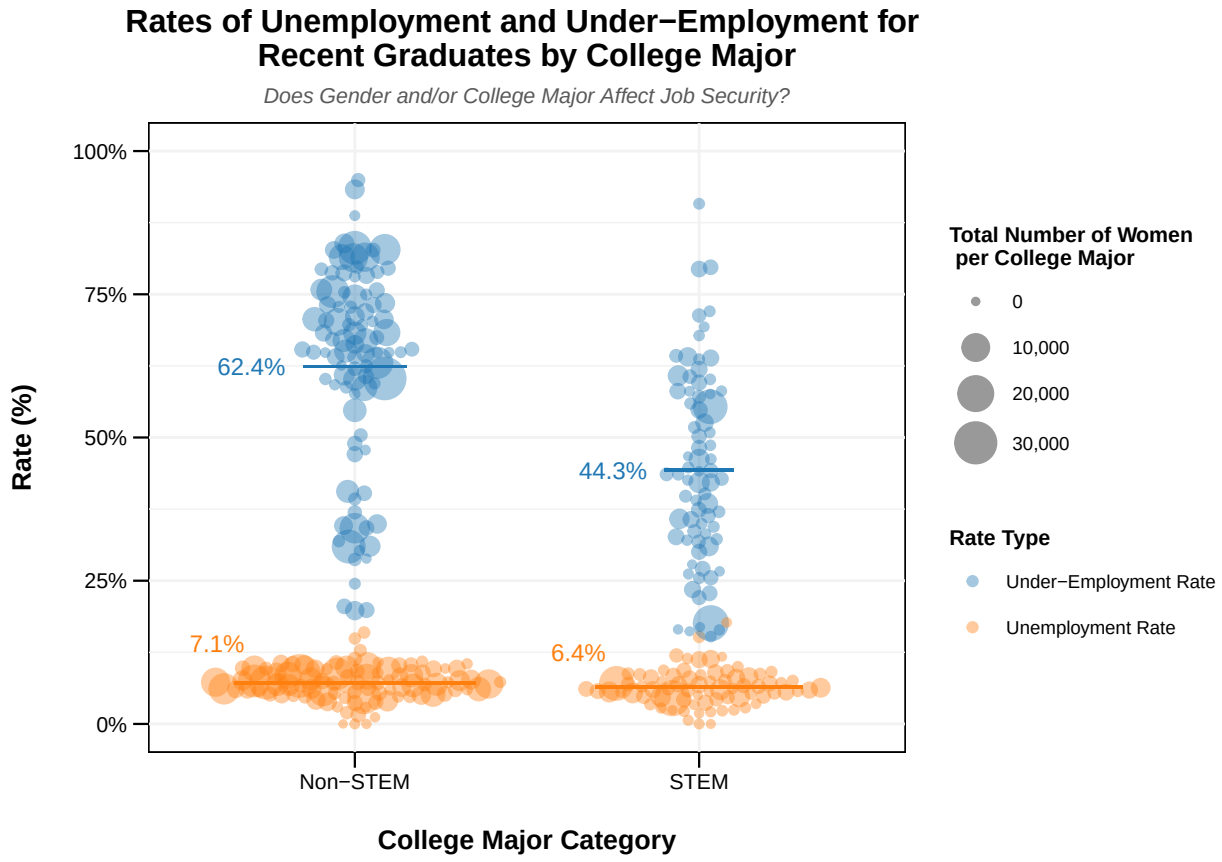
What Majors have the greatest job security and does a Gender bias exist with respect to employment?

Figure 5

```
data.un.1 <- recent.grads %>% mutate(Underemploy_Rate = (Low_wage_jobs +
    Non_college_jobs)/(Low_wage_jobs + Non_college_jobs +
    College_jobs)) %>% mutate(Cat2 = ifelse(Major_category %in%
    women.stem$Major_category, "STEM", "Non-STEM")) %>%
    pivot_longer(cols = c("Underemploy_Rate", "Unemployment_rate"),
        names_to = "Type", values_to = "Rate") %>%
    select(Rank, Median, Women, Major, Major_category,
        Type, Rate, ShareWomen) %>% mutate(Cat2 = ifelse(Major_category %in%
```

```
    women.stem$Major_category, "STEM", "Non-STEM"),
    ModSal = 3 * Median) %>% arrange(desc(Rank))

# rerank major category
data.un.1$Major_category <- factor(data.un.1$Major_category,
    levels = popdata$Major_category)

# Remove NA's
data.un.1 <- na.omit(data.un.1)

# ggbeeswarm library
library(ggbeeswarm)
library(ggsci)

# Summary Stats

sum1 <- data.un.1 %>% filter(Type == "Unemployment_rate" &
    Cat2 == "STEM")
sum2 <- data.un.1 %>% filter(Type == "Unemployment_rate" &
    Cat2 == "Non-STEM")
sum3 <- data.un.1 %>% filter(Type == "Underemploy_Rate" &
    Cat2 == "STEM")
sum4 <- data.un.1 %>% filter(Type == "Underemploy_Rate" &
    Cat2 == "Non-STEM")

sum1 <- mean(sum1$Rate)
sum2 <- mean(sum2$Rate)
sum3 <- mean(sum3$Rate)
sum4 <- mean(sum4$Rate)

scipen = 999

# Unemployment Plot
ggplot(data.un.1) + geom_beeswarm(aes(x = Rate, y = Cat2,
    col = Type, size = Women), cex = 1.7, alpha = 0.4,
    groupOnX = FALSE) + scale_size_continuous(range = c(1,
    7), labels = c("0", "10,000", "20,000", "30,000")) +
    scale_color_d3(labels = c("Under-Employment Rate",
        "Unemployment Rate")) + scale_fill_d3() + theme_minimal() +
    scale_x_continuous(label = percent_format(), limits = c(0,
        1), breaks = seq(0, 1, 0.25)) + coord_flip() +
    geom_segment(x = sum2, xend = sum2, y = 0.65, yend = 1.35,
        color = "#FF7F0EFF", size = 0.5) + geom_segment(x = sum4,
    xend = sum4, y = 0.85, yend = 1.15, color = "#1F77B4FF",
    size = 0.5) + geom_segment(x = sum1, xend = sum1,
    y = 1.7, yend = 2.3, color = "#FF7F0EFF", size = 0.5) +
    geom_segment(x = sum3, xend = sum3, y = 1.9, yend = 2.1,
        color = "#1F77B4FF", size = 0.5) + annotate("text",
    x = sum2 + 0.07, y = 0.6, size = 3, color = "#FF7F0EFF",
    label = paste0(round(sum2 * 100, digits = 1), "%")) +
    annotate("text", x = sum4, y = 0.7, size = 3, color = "#1F77B4FF",
        label = paste0(round(sum4 * 100, digits = 1),
            "%")) + annotate("text", x = sum1 + 0.06,
```

```
    y = 1.65, size = 3, color = "#FF7F0EFF", label = paste0(round(sum1 *
        100, digits = 1), "%")) + annotate("text",
    x = sum3, y = 1.75, size = 3, color = "#1F77B4FF",
    label = paste0(round(sum3 * 100, digits = 1), "%")) +
    theme(panel.grid.major = element_line(color = "gray95"),
        panel.grid.minor = element_line(color = "gray95"),
        panel.background = element_rect(fill = NA),
        plot.title = element_text(size = 12, face = "bold",
            color = "black", hjust = 0.5), legend.title = element_text(size = 8,
            color = "black", face = "bold"), axis.text = element_text(size = 8,
            color = "black", hjust = 0.5), axis.ticks = element_line(size = 0.4),
        axis.ticks.length = unit(0.2, "cm"), panel.border = element_rect(fill = NA,
            size = 0.25, color = "black"), legend.text = element_text(size = 7,
            color = "black"), axis.title = element_text(size = 10,
            face = "bold", color = "black"), plot.subtitle = element_text(size = 8,
            color = "grey35", hjust = 0.5, face = "italic"),
        plot.caption = element_text(size = 5, color = "grey60",
            hjust = 0), legend.box.just = "left", legend.direction = "vertical",
        legend.title.align = 0, legend.key.size = unit(1.2,
            "lines")) + labs(title = "Rates of Unemployment and Under-Employment for \nRecent Graduates
    subtitle = "Does Gender and/or College Major Affect Job Security?",
    x = "Rate (%)\n", y = "\nCollege Major Category\n",
    color = "Rate Type", size = "Total Number of Women\n per College Major",
    caption = "Data Source : https://github.com/fivethirtyeight/data/blob/master/college-majors/recent-g
Note1: Underemployment is defined as the percentage of respondents currently employed in jobs not requi
Note2: STEM Majors include those in Science, Technology, Engineering & Math")
```



Executive Summary

The preceding data visualization project principally examined the relevant dataset with the goal of providing the reader key information to guide college major choice. For most, the choice of college major soon after matriculation is a hurried process based upon little research and background information. The visualizations provided intially looked at two metrics that may guide one's choice of college major: (1) popularity and (2) starting salary potential.

The analysis indicated that generally the most popular college majors are the most lucrative. For instance, popular majors such as Engineering and Business tended to afford the highest starting salaries. However, when broken down by gender participation this relationship falls noticeably short. College majors that are dominated by female participation (Humanites and Liberal Arts) tended to be the least lucrative. This beckons the potentially controversial question: are females purposefully self-selecting fields that inherently lower in pay or does society inherently undervalue the work product of female-dominated professions?

The visualizations explored the potential earnings power for those graduating with STEM (Science Technology, Engineering and Math) degrees. Not surprisingly, those graduating with these degrees tended to have higher starting salaries. Interestingly, the gender bias of starting salary was also highly evident for STEM degree recipients. The most lucrative STEM degree, Engineering, was male dominated. While, the least lucrative STEM degrees, Health/Biology, were female dominated. The accepted reasoning for this is the concept of valuative bias [1], where society unfairly characterizes female-dominated professions as less valuable.

Fortunately, the last visualization indicates that gender bias is not readily apparent with respect to unemployment ar under-employment. Instead, the influence of college major choice has a profound influence on under and un-employment, with those possessing STEM degrees having lower rates.

In general, the visualizations provide enhanced information to guide college major choice and highlight a profound gender bias that female graduates suffer. Hopefully the visualizations will inspire change that will result in salary parity amongst genders.

[1] https://www.naceweb.org/job-market/compensation/exploring-gender-bias-in-starting-salary-offers-among-stem-majors/

Shiny Dashboard

Exploring Gender Bias in Starting Salaries and Unemployment

The reader is provided a URL below to an RShiny dashboard that affords an interactive view of the relevant dataset. The dashboard is designed to not only explore the economic ramifications of college major choice, but will also provide a means to explore the influence of gender on both starting salaries and employment.

[linked phrase] (https://owen-evans.shinyapps.io/EvansOwen_DSA5200_FinalProject/)

END

```
#
#setwd("~/Desktop/Final Project/Shiny Dashboard")
library(shiny)
library(shinydashboard)
library(plotly)

shinyUI(
    dashboardPage(title="Gender Bias",

        dashboardHeader(title=span("Exploring Gender Bias in Starting Salaries and Unemployment",
                                   style = "color: white; font-size: 28px;"), titleWidth = '99%'),

        dashboardSidebar(
            sidebarMenu(width=100,
                 menuItem(h4("Starting Salary"), tabName= "median"),
                 menuItem(h4("Unemployment Rate"), tabName = "unemploy"),
                 menuItem(h4("Under-Employment Rate"), tabName = "underemploy")
                            ) #sidebar menu
                        ), #dashboardsidebar
        dashboardBody(
            tags$head(tags$style(HTML('
                .skin-blue .main-header .logo {
                  background-color: #3c8dbc;
                }
                .skin-blue .main-header .logo:hover {
                  background-color: #3c8dbc;
                }'))),
                      tabItems(
                          tabItem(tabName = "median", status="primary",
                                 fluidRow(
                                     column(12, align="center",
```

```
                                h3(strong("Choosing a Major - Economic Ramifications & Gende
                                h4("The following dashboard utilizes data from a American Co
                                   Survey 2010-2012 Public Use Microdata Series to explore t
                                   college major choice on starting salary and empolyment.
                                   Of particular interest is strong evidence of gender bias
                                   starting salaries and choice of college major.
                                   The reader is encouraged to explore these relationships t
                                   interactive visualizations below.")
                                )
                            ),

                            fluidRow(
                               column(8,
                                   h3(strong("Starting Salary")),
                                   box(plotlyOutput("salaryplot"),
                                   status="primary", width=12)), # End Column
                               column(4,
                                   fluidRow(
                                       column(12,
                                           h3(strong("Plot Filters")))), #end of row
                                   box(status = "primary", size=12,
                                       fluidRow(
                                       column(12,
                                           selectInput("stem", label="Major Category",
                                           c(NULL, "STEM Major"="STEM", "non-STEM Major"="non-STEM
                                           selected=NULL))
                                               ) # end column
                                           ), #end row
                                       fluidRow(
                                           column(12,
                                               selectInput(inputId = "majorcat",
                                               label="Major Sub-Category",
                                               choices= "",
                                               selected="")
                                               ) # end column
                                           ),# end row
                                       ) #endbox
                                   ) # end short column
                               ),
                               fluidRow(
                                   column(12,align="left",
                                          tags$i("A Shinydashboard by Owen R. Evans, DSA5200 Final Proj
                                          tags$br(""),
                                          tags$a(href="https://github.com/fivethirtyeight/data/tree/mas
                                                 "Data Source")
                                       ))# End row
                           ), # end tab item

                   tabItem(tabName = "unemploy", status="primary",
                           fluidRow(
                               column(12, align="center",
                                      h3(strong("Choosing a Major - Economic Ramifications & Gender I
                                      h4("The following dashboard utilizes data from a American Commu
```

```
                                Survey 2010-2012 Public Use Microdata Series to explore t
                                college major choice on starting salary and empolyment.
                                Of particular interest is strong evidence of gender bias
                                starting salaries and choice of college major.
                                The reader is encouraged to explore these relationships i
                                interactive visualizations below.")
                        )
                    ),
                      fluidRow(
                            column(8,
                                   h3(strong("Unemployment Rate")),
                                   box(plotlyOutput("unemployplot"),
                                       status="primary", width=12)), # End Column
                            column(4,
                                   fluidRow(
                                       column(12,
                                              h3(strong("Plot Filters")))),# end row
                                 box(status = "primary", size=12,
                                   fluidRow(
                                       column(12,
                                              selectInput("stem2", label="Major Category",
                                                          c(NULL, "STEM Major"="STEM", "non-ST
                                                            selected=NULL))
                                       ) # end column
                                   ), #end row
                                   fluidRow(
                                       column(12,
                                              selectInput(inputId = "majorcat2",
                                                          label="Major Sub-Category",
                                                          choices= "",
                                                          selected="")
                                       ) # end column
                                   ) # end row
                                )# end box
                            ) # end short column
                        ),
                    fluidRow(
                        column(12,align="left",
                               tags$i("A Shinydashboard by Owen R. Evans, DSA5200 Final Proje
                               tags$br(""),
                               tags$a(href="https://github.com/fivethirtyeight/data/tree/maste
                                      "Data Source")
                        ))# End bottom row #fluidrow
                ),  #tabitem-unemploy
            tabItem(tabName = "underemploy", status = "primary",
                    fluidRow(
                        column(12, align="center",
                               h3(strong("Choosing a Major - Economic Ramifications & Gend
                               h4("The following dashboard utilizes data from a American C
                                 Survey 2010-2012 Public Use Microdata Series to explore t
                                 college major choice on starting salary and empolyment.
                                 Of particular interest is strong evidence of gender bias
                                 starting salaries and choice of college major.
```

```
                                        The reader is encouraged to explore these relationships i
                                        interactive visualizations below.")
                                    )
                                ),
                                fluidRow(
                                    column(8,
                                           h3(strong("Under-employment Rate")),
                                           box(plotlyOutput("underplot"),
                                               status="primary", width=12)), # End Column
                                    column(4,
                                           fluidRow(
                                               column(12,
                                                      h3(strong("Plot Filters")))), #end row
                                           box(status = "primary", size=12,
                                           fluidRow(
                                               column(12,
                                                      selectInput("stem3", label="Major Category",
                                                       c(NULL, "STEM Major"="STEM", "non-STEM Major"=
                                                                    selected=NULL))
                                               ) # end column
                                           ), #end row
                                           fluidRow(
                                               column(12,
                                                      selectInput(inputId = "majorcat3",
                                                                  label="Major Sub-Category",
                                                                  choices= "",
                                                                  selected="")
                                               ) # end column
                                           ) # end row
                                       ) # end box
                                    ) # end short column
                                ), # end of row
                                fluidRow(
                                    column(12,align="left",
                                           tags$i("A Shinydashboard by Owen R. Evans, DSA5200 Final P
                                           tags$br(""),
                                           tags$a(href="https://github.com/fivethirtyeight/data/tree/
                                                  "Data Source")
                                    ))
                            ) #end of tab item
                        )#tabitems
                    )#dashboardbody
                )#dashboardpage
    )#shinyUI
```

```
library(shiny)
library(shinydashboard)
library(ggplot2)
library(dplyr)
library(plotly)
library(scales)
library(ggsci)
library(stringr)
```

```r
# Load College Major Data women.stem <-
# read.csv('women.stem.csv')

women.stem <- read.csv("womenstem1.csv")
# read.csv(paste0('https://raw.githubusercontent.com/fivethirtyeight',
#'/data/master/college-majors/women-stem.csv'))  # women in STEM degrees

recent.grads <- read.csv("recentgrads1.csv")
# read.csv(paste0('https://raw.githubusercontent.com/fivethirtyeight',
#'/data/master/college-majors/recent-grads.csv')) # All data for recent graduates


# Data Preparation
recent.grads <- na.omit(recent.grads)
recent.grads$Major <- str_to_title(recent.grads$Major)
recent.grads <- recent.grads %>% mutate(Cat1 = ifelse(Major_category %in%
    women.stem$Major_category, "STEM", "non-STEM")) %>%
    mutate(under = ((Non_college_jobs + Low_wage_jobs)/(Non_college_jobs +
        Low_wage_jobs + College_jobs))) %>% mutate(popup1 = paste0("Major: ",
    Major, "<br>", "Major Sub-Category: ", Major_category,
    "<br>", "Percent Female: ", round(ShareWomen *
        100, digits = 2), "%", "<br>", "Median Salary: ",
    "$", Median, "<br>")) %>% mutate(popup2 = paste0("Major: ",
    Major, "<br>", "Major Sub-Category: ", Major_category,
    "<br>", "Percent Female: ", round(ShareWomen *
        100, digits = 2), "%", "<br>", "Unemployment Rate: ",
    round(Unemployment_rate * 100, digits = 2), "%",
    "<br>")) %>% mutate(popup3 = paste0("Major: ",
    Major, "<br>", "Major Sub-Category: ", Major_category,
    "<br>", "Percent Female: ", round(ShareWomen *
        100, digits = 2), "%", "<br>", "Under-Employment Rate: ",
    round(under * 100, digits = 2)))

x <- which(recent.grads$Major_category %in% women.stem$Major_category)
nonstem.data <- recent.grads[-x, ]
stem.data <- recent.grads[x, ]

# Server Logic
shinyServer(function(session, input, output) {

    # Change sub category input based upon category
    # selection
    observe({
        x <- recent.grads %>% filter(Cat1 == input$stem) %>%
            select(Major_category) %>% rename(`Major Sub-Categories` = Major_category)
        y <- paste0("All ", input$stem, " Majors")
        updateSelectInput(session, "majorcat", "Major Sub-Categories",
            choices = c(y, unique(x)))
    })  #observe event, changing submenu

    observe({
        x <- recent.grads %>% filter(Cat1 == input$stem2) %>%
            select(Major_category)
```

```
    y <- paste0("All ", input$stem2, " Majors")
    updateSelectInput(session, "majorcat2", "Major Sub-Categories",
        choices = c(y, unique(x)))
})  #observe event2, changing submenu

observe({
    x <- recent.grads %>% filter(Cat1 == input$stem3) %>%
        select(Major_category)
    y <- paste0("All ", input$stem3, " Majors")
    updateSelectInput(session, "majorcat3", "Major Sub-Categories",
        choices = c(y, unique(x)))
})  #observe event3, changing submenu

# Filter data based on input
plotdata.sal <- reactive({
    if (input$majorcat %in% recent.grads$Major_category) {
        filter.data <- recent.grads %>% filter(Major_category ==
            input$majorcat)
    } else if (input$stem == "STEM") {
        filter.data <- recent.grads %>% filter(Cat1 ==
            "STEM")
    } else if (input$stem == "non-STEM") {
        filter.data <- recent.grads %>% filter(Cat1 ==
            "non-STEM")
    }
    return(filter.data)
})

# Salary Plot Render
output$salaryplot <- renderPlotly({
    p <- ggplot(plotdata.sal(), aes(x = ShareWomen,
        y = Median, text = popup1, group = 1)) +
        geom_point(aes(color = Major_category),
            width = 0.3, alpha = 0.8) + geom_smooth(se = FALSE,
        color = "black", size = 0.5) + scale_fill_d3() +
        scale_x_continuous(labels = label_percent(),
            limits = c(0, 1)) + scale_y_continuous(labels = label_dollar()) +
        theme_bw() + theme(legend.title = element_blank(),
        panel.grid.major = element_line(color = "gray85"),
        panel.grid.minor = element_line(color = "gray65"),
        axis.title = element_text(face = "bold",
            size = 12, color = "black"), axis.text = element_text(size = 10,
            color = "black"), panel.background = element_rect(size = 0.25,
            color = "black"), legend.text = element_text(size = 8)) +
        labs(x = "\nPercent Female per Major Category",
            y = "Starting Salary \n", caption = paste0("Data Source:   https://github.com",
              "/fivethirtyeight/data/tree/master/",
              "college-majors"))

    plot <- ggplotly(p, tooltip = "popup1")

    for (i in 1:length(plot$x$data)) {
        if (!is.null(plot$x$data[[i]]$name))
```

```
            {
              plot$x$data[[i]]$name = gsub("\\(",
                "", str_split(plot$x$data[[i]]$name,
                  ",")[[1]][1])
            }  # Plotly Legend 'Bug'
    }  # end for loop
    plot
})  # End Sal Plot

plotdata.unemploy <- reactive({
    if (input$majorcat2 %in% recent.grads$Major_category) {
        filter.data <- recent.grads %>% filter(Major_category ==
            input$majorcat2)
    } else if (input$stem2 == "STEM") {
        filter.data <- recent.grads %>% filter(Cat1 ==
            "STEM")
    } else if (input$stem2 == "non-STEM") {
        filter.data <- recent.grads %>% filter(Cat1 ==
            "non-STEM")
    }
    return(filter.data)
})

output$unemployplot <- renderPlotly({
    p <- ggplot(plotdata.unemploy(), aes(x = ShareWomen,
        y = Unemployment_rate, text = popup2, group = 1)) +
        geom_point(aes(color = Major_category),
            width = 0.3, alpha = 0.8) + geom_smooth(se = FALSE,
        color = "black", size = 0.5) + scale_fill_d3() +
        scale_x_continuous(labels = label_percent()) +
        scale_y_continuous(labels = label_percent()) +
        theme_bw() + theme(legend.title = element_blank(),
        panel.grid.major = element_line(color = "gray85"),
        panel.grid.minor = element_line(color = "gray65"),
        axis.title = element_text(face = "bold",
            size = 12), axis.text = element_text(size = 10),
        legend.text = element_text(size = 8)) +
        labs(x = "\nPercent Female per Major Category",
            y = "Unemployment Rate \n", caption = paste0("Data Source:   https://github.com",
              "/fivethirtyeight/data/tree/master/",
              "college-majors"))

    plot <- ggplotly(p, tooltip = "popup2")

    for (i in 1:length(plot$x$data)) {
        if (!is.null(plot$x$data[[i]]$name))
            {
              plot$x$data[[i]]$name = gsub("\\(",
                "", str_split(plot$x$data[[i]]$name,
                  ",")[[1]][1])
            }  # Plotly Legend 'Bug'
    }  # end for loop
    plot
```

```
    })  # End Unemploy Plot

    plotdata.under <- reactive({
        if (input$majorcat3 %in% recent.grads$Major_category) {
            filter.data <- recent.grads %>% filter(Major_category ==
                input$majorcat3)
        } else if (input$stem3 == "STEM") {
            filter.data <- recent.grads %>% filter(Cat1 ==
                "STEM")
        } else if (input$stem3 == "non-STEM") {
            filter.data <- recent.grads %>% filter(Cat1 ==
                "non-STEM")
        }
        return(filter.data)
    })

    output$underplot <- renderPlotly({
        p <- ggplot(plotdata.under(), aes(x = ShareWomen,
            y = under, text = popup3, group = 1)) +
            geom_point(aes(color = Major_category),
                width = 0.3, alpha = 0.8) + geom_smooth(se = FALSE,
            color = "black", size = 0.5) + scale_fill_d3() +
            scale_x_continuous(labels = label_percent()) +
            scale_y_continuous(labels = label_percent()) +
            theme_bw() + theme(legend.title = element_blank(),
            panel.grid.major = element_line(color = "gray85"),
            panel.grid.minor = element_line(color = "gray65"),
            axis.title = element_text(face = "bold",
                size = 12), axis.text = element_text(size = 10),
            legend.text = element_text(size = 8)) +
            labs(x = "\nPercent Female per Major Category",
                y = "Under-Employment Rate \n", caption = paste0("Data Source:   https://github.com",
                  "/fivethirtyeight/data/tree/master/",
                  "college-majors"))

        plot <- ggplotly(p, tooltip = "popup3")

        for (i in 1:length(plot$x$data)) {
            if (!is.null(plot$x$data[[i]]$name))
                {
                  plot$x$data[[i]]$name = gsub("\\(",
                    "", str_split(plot$x$data[[i]]$name,
                      ",")[[1]][1])
                }  # Plotly Legend 'Bug'
        }  # end for loop
        plot
    })  # End Sal Plot

})  # end server
```