**PSYC 3031 A**

**Monique Herbert**

**Data Visualization Project**

Francis Zita 216642217

Alon Cohen 2005904

Eros Garcia 215516081

**November 4th, 2020**

**Description of Data:**

The data selected for this visualization is a cross sectional survey conducted by Psychology Today, in 1969.  The data included 601 participants with nine different variables.  The first variable is ‘affairs’, which is categorical.  The question asked in the survey was how often an individual engaged in extramarital sexual intercourse in the last year.  This was categorized as follows: ‘0 = none’, ‘1 = once’, ‘2 = twice’, 3 = three times’, ‘7 = four to ten times’, ‘12 = monthly, weekly, and daily’.  The next variable is ‘gender’, which was categorized as female or male.  The third variable is ‘age’ in years, which was categorized as: ‘17.5 = under 20’, ‘22 = 20–24’, ‘27 = 25–29’, ‘32 = 30–34’, ‘37 = 35–39’, ‘42 = 40–44’, ‘47 = 45–49’, ‘52 = 50–54’, ‘57 = 55 or over’.  The fourth variable is ‘years married’, which was categorized as: ‘0.125 = 3 months or less’, ‘0.417 = 4–6 months’, ‘0.75 = 6 months–1 year’, ‘1.5 = 1–2 years’, ‘4 = 3–5 years’, ‘7 = 6–8 years’, ‘10 = 9–11 years’, ‘15 = 12 or more years’.  The fifth variable is ‘children’.  The questionnaire asked individuals to either answer ‘yes’ or ‘no’ if the couple currently had children.  The sixth variable is ‘religiousness’.  Participants were asked to rate how religious they currently are.  This was categorized as: ‘1 = anti’, ‘2 = not at all’, ‘3 = slightly’, ‘4 = somewhat’, ‘5 = very’.  The seventh variable is ‘education’.  The questionnaire asked participants to identify what level of education they currently have.  This was categorized as: ‘9 = grade school’, ‘12 = high school graduate’, ‘14 = some college’, ‘16 = college graduate’, ‘17 = some graduate work’, ‘18 = master's degree’, ‘20 = Ph.D., M.D., or other advanced degree’.  The eighth variable is ‘occupation’.  Participants were asked to pick a number indicating their occupation according to Hollingshead’s classification in reverse numbering.  The last variable is ‘rating’.  The questionnaire asked participants to rate their happiness in the marriage.  This was categorized as: ‘1 = very unhappy’, ‘2 = somewhat unhappy’, ‘3 = average’, ‘4 = happier than average’, ‘5 = very happy’.

**Visualization Question:**

Our visualization question is whether gender plays a role in the likelihood of people having an affair, and how this relationship changes with each participant’s religiousness and rating of marriage.

**Description of Goals/Outcomes:**

Given the focus of the visualization question, the data was changed to show specific factors.  Instead of nine variables, our visualization will only use four variables, which will be recoded.  The first variable will be ‘gender’, which is either male or female.  The second variable will be ‘affairs’.  This was changed to identify if an individual engaged in an affair at least once in the past year, instead of looking at the overall number of affairs.  The categories are now: 0 = faithful, 1,2,3,7, and 12 = cheated.  The third variable will be ‘religiousness’.  Since the visualization will focus on whether a participant is religious or not, the categories are now: 1 = Not Religious, 2 = Not Religious, 3 = Religious, 4 = Religious, and 5 = Religious.  The final variable will be ‘rating’.  This was changed in order to identify whether individuals were happy or not in their marriage.  This was recoded as: 1 = Unhappy, 2= Unhappy, 3 = Neutral, 4 = Happy, 5 = Happy.

The goal and outcome of this visualization is to demonstrate any gender differences in extramarital affairs.  This will be shown using a stacked bar graph, which will allow us to see any percentage difference (Figure 1).  In this graph, there is no significant difference between females and males and affairs.

Chart

Description automatically generated

*Figure 1*: Faithfulness between genders

Next, we will show how the relationship between gender and affairs changes when religiousness is added as a third variable.  We show this by using a stacked bar graph (Figure 2).  When religiousness was added as a third variable in the graph, we see a considerable increase in affairs for non-religious males.

**Chart, bar chart

Description automatically generated**

*Figure 2*: Faithfulness between genders and religious orientations

Finally, we will show how the relationship between gender and affairs changes when rating of the marriage is added as a third variable.  We will display this change by using a stacked bar graph (Figure 3).  When the rating of marriage was added as a third variable, this graph shows a reverse trend between neutral males and females, with females having more affairs.  We also see a significant increase in affairs by both males and females who were unhappy.

**Chart, bar chart

Description automatically generated**

*Figure 3*: Faithfulness between genders and marriage satisfaction

**Limitations:**

A limitation of the data we used is that it does not include all the variables that were collected in the survey.  This was done to limit the scope of our visualization so we could focus on our research question.  Another limitation is that the data was collected through Psychology Today.  People who didn’t have access to Psychology Today would be excluded from the sample, and therefore limit representation.  Another limitation in the data used is that current categories of gender have expanded beyond typical male or female options.  This limits the generalizability to populations who identify in newer classifications.  The last limitation of the data is that it was collected in 1969 and would not be applicable to the present population.

**R Code and Output:**

**Group-Project.R**

2020-11-04

# Group Projects: Gender and Extramarital Affairs  
  
# Access the needed libraries:  
  
# The purpose of the tidyverse package is that it contains the graphs and other necessary packages that we need for our code.   
  
library(tidyverse)

## -- Attaching packages ----------------------------------------------------------------------------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2 v purrr 0.3.4  
## v tibble 3.0.3 v dplyr 1.0.2  
## v tidyr 1.1.2 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0

## -- Conflicts -------------------------------------------------------------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

# ghibli contains different colours for the graphs that we are using  
  
library(ghibli)

## Warning: package 'ghibli' was built under R version 4.0.3

# Load data  
# Here, we are loading the data file "Affairs.csv", to a data frame called "affairs"   
# After running "read\_csv", it shows us the types of data that we have from the data file; for example we get "col\_double" for some data, and "col\_character" for the others.  
  
affairs <- read\_csv("Data/Affairs.csv")

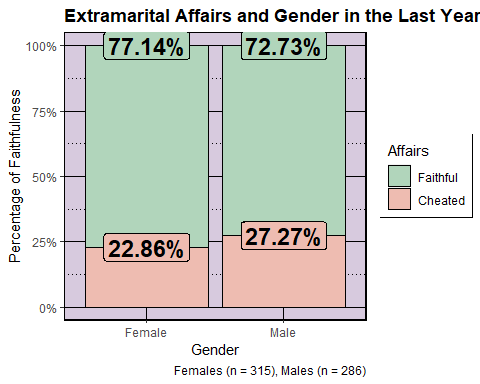
## Warning: Missing column names filled in: 'X1' [1]

## Parsed with column specification:  
## cols(  
## X1 = col\_double(),  
## affairs = col\_double(),  
## gender = col\_character(),  
## age = col\_double(),  
## yearsmarried = col\_double(),  
## children = col\_character(),  
## religiousness = col\_double(),  
## education = col\_double(),  
## occupation = col\_double(),  
## rating = col\_double()  
## )

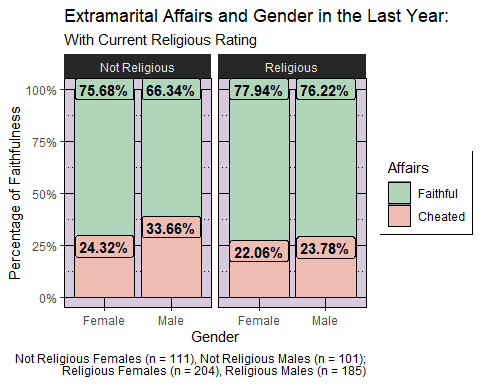
# The next line confirms that there are as many distinct IDs as there are rows  
  
count(distinct(affairs, X1)) == count(affairs)

## n  
## [1,] TRUE

# Change codes to appropriate descriptions:  
# Here we are mutating certain information from the "affairs" data frame, turning numerical data into character data before recoding it, and creating a new data frame called "affairsNew".  
# First, we are recoding the rating column from numbers to factors - "unhappy" (1-2); "neutral" (3); "happy" (4-5)   
  
affairsNew <- mutate(affairs,  
 rating = as.character(rating),  
 rating = fct\_recode(rating,  
 "Unhappy" = "1",  
 "Unhappy" = "2",  
 "Neutral" = "3",  
 "Happy" = "4",  
 "Happy" = "5")) %>%  
   
   
 # Next, we are recoding the religiousness column - "not religious" (1-2); "religious" (3-5,)  
 # We are only interested in knowing if they are religious or not.  
   
 mutate(religiousness = as.character(religiousness),  
 religiousness = fct\_recode(religiousness,  
   
 'Not Religious' = "1",  
 'Not Religious' = "2",  
 'Religious' = "3",  
 'Religious' = "4",  
 'Religious' = "5")) %>%  
   
   
   
 # Finally, we are recoding the affairs column - "cheated" (1,2,3,7,12); "faithful" (0)  
   
 mutate(affairs = as.character(affairs),  
 affairs = fct\_recode(affairs,  
 'Cheated' = "1",  
 'Cheated' = "2",  
 'Cheated' = "3",  
 'Cheated' = "7",  
 'Cheated' = "12",  
 'Faithful' = "0")) %>%  
   
 # The next lines changes first letters from lowercase to uppercase  
   
 mutate(gender = fct\_recode(gender,  
 'Male' = 'male',  
 'Female' = 'female')) %>%  
   
 rename(Affairs = affairs)  
  
# Here we are creating a new data frame where only the rows with people who cheated will be added  
  
cheated <- filter(affairsNew, Affairs == "Cheated")   
  
  
# Here we are creating a new data frame where only the rows with people who are faithful will be added  
  
faithful <- filter(affairsNew, Affairs == "Faithful")  
  
# Here we are creating a new data frame where only the rows with people who are religious will be added  
  
religious <- filter(affairsNew, religiousness == "Religious")  
  
# Here we are creating a new data frame where only the rows with people who are not religious will be added  
  
notReligious <- filter(affairsNew, religiousness != "Religious")  
  
# Here we are creating a new data frame where only the rows with people who are happy with their marriage will be added  
  
happy <- filter(affairsNew, rating == "Happy")  
  
# Here we are creating a new data frame where only the rows with people who are neutral will be added  
  
neutral <- filter(affairsNew, rating == "Neutral")  
  
# Here we are creating a new data frame where only the rows with people who are unhappy will be added  
  
unhappy <- filter(affairsNew, rating == "Unhappy")  
  
# The upcoming line creates a vector with two values of each for the total number of Females and Males  
  
totalByGender <- c(rep(sum(affairsNew$gender == "Female"),2),  
 rep(sum(affairsNew$gender == "Male"),2))  
  
  
# Here we are creating a graph from the data frame "affairsNew" comparing gender which is in the x-axis, by affairs which is in the y-axis  
# geom\_bar(position = "fill") makes it a stacked bar graph  
  
ggplot(affairsNew,   
 aes(x = gender, fill = Affairs)) +  
 geom\_bar(position = "fill", colour="black") +  
   
 # The next line changes the color of the graph  
   
 scale\_fill\_manual(values = c("#B1D5BBFF","#EEBCB1FF")) +  
   
 # The labs() function adds a bold title, labels to the x-axis and y-axis of the graph, as well as a caption  
   
 labs(title = "Extramarital Affairs and Gender in the Last Year:",  
 caption = "Females (n = 315), Males (n = 286)",  
 x = "Gender",  
 y = "Percentage of Faithfulness") +  
 theme(plot.title = element\_text(face = "bold")) +  
   
 # The next few lines uses a theme dark template and overrides with grid design and legend outline  
   
 theme(legend.box.background = element\_rect(colour = "black")) +  
 theme(  
 panel.background = element\_rect(fill = "#D7CADEFF",  
 colour = "black",  
 size = 1, linetype = "solid"),  
 panel.grid.major = element\_line(size = 0.5, linetype = 'solid',  
 colour = "black"),  
 panel.grid.minor = element\_line(size = 0.5, linetype = 'dotted',  
 colour = "black")  
 ) +  
   
 # The geom\_text function adds the value of percentages to each proportion of the bars; from "%0.2f%%", we print it with two decimals precision, calculating it using the count for each category of faithfulness, divided by the vector of total males and females  
   
 geom\_label(aes(label=sprintf("%0.2f%%", 100\*stat(count)/totalByGender)), stat='count', position='fill',  
 fontface = "bold", size = 6, show.legend = FALSE) +  
   
 # the next line turns the y-axis into a percentage instead of a ratio  
   
 scale\_y\_continuous(labels = scales::percent)



# The upcoming lines creates a vector named "totalsReligious" with values for the total number of non religious females and males, and religious females and males.  
  
totalsReligious <- c(rep(sum(notReligious$gender == "Female"), 2),  
 rep(sum(notReligious$gender == "Male"), 2),  
 rep(sum(religious$gender == "Female"), 2),  
 rep(sum(religious$gender == "Male"), 2))  
  
# Here we are creating a stacked bar graph that compares religious and non religious males and females on the percentage of their faithfulness.  
  
ggplot(affairsNew,  
 aes(x = gender, fill = Affairs)) +  
 geom\_bar(position = "fill",colour="black" ) +  
   
 # The next line changes the color of the graph again  
   
 scale\_fill\_manual(values = c("#B1D5BBFF","#EEBCB1FF")) +  
   
   
 # The labs() function adds a bold title, labels to the x-axis and y-axis of the graph, as well as a caption and a subtitle  
   
 labs(title = "Extramarital Affairs and Gender in the Last Year:",  
 subtitle = "With Current Religious Rating",  
 caption = "Not Religious Females (n = 111), Not Religious Males (n = 101);\nReligious Females (n = 204), Religious Males (n = 185)",  
 x = "Gender",  
 y = "Percentage of Faithfulness") +  
 theme(plot.title = element\_text(face = "bold")) +  
   
 # The next few lines uses a theme dark template and overrides with grid design and legend outline  
   
 theme\_dark() +  
 theme(legend.box.background = element\_rect(colour = "black")) +  
 theme(  
 panel.background = element\_rect(fill = "#D7CADEFF",  
 colour = "black",  
 size = 0.5, linetype = 1),  
 panel.grid.major = element\_line(size = 0.5, linetype = 'solid',  
 colour = "black"),  
 panel.grid.minor = element\_line(size = 0.5, linetype = 'dotted',  
 colour = "black")  
 ) +  
   
 # The geom\_text function adds the value of percentages to each proportion of the bars; from "%0.2f%%", we print it with two decimals precision, calculating it using the count for each category of faithfulness, divided by the vector of total nonreligious males and females, and total religious males and females  
   
 geom\_label(aes(label=sprintf("%0.2f%%", 100\*stat(count)/totalsReligious)), stat='count', position='fill', fontface = "bold", size = 4, show.legend = FALSE) +  
   
 # the next line turns the y-axis into a percentage instead of a ratio  
   
 scale\_y\_continuous(labels = scales::percent) +  
   
 # the next line makes the split in the graphs between Not Religious and Religious  
   
 facet\_grid(~religiousness)



# The upcoming lines creates a vector named "totalsUnhappy" with values for the total number of unhappy, neutral, and happy females and males.  
  
totalsUnhappy <- c(rep(sum(unhappy$gender == "Female"), 2),  
 rep(sum(unhappy$gender == "Male"), 2),  
 rep(sum(neutral$gender == "Female"), 2),  
 rep(sum(neutral$gender == "Male"), 2),  
 rep(sum(happy$gender == "Female"), 2),  
 rep(sum(happy$gender == "Male"), 2))   
  
# Here we are creating a stacked bar graph that compares unhappy, neutral, and happy females and males on the percentage of their faithfulness.  
  
ggplot(data = affairsNew,  
 mapping = aes(x = gender, fill = Affairs)) +  
 geom\_bar(position = "fill", colour="black") +  
   
 # The next line changes the color of the last graph  
 scale\_fill\_manual(values = c("#B1D5BBFF","#EEBCB1FF")) +  
   
 # The labs() function adds a bold title, labels to the x-axis and y-axis of the graph, as well as a caption and a subtitle  
   
 labs(title = "Extramarital Affairs and Gender in the Last Year:",  
 subtitle = "With Marriage Satisfaction",  
 caption = "Unhappy Females (n = 46), Unhappy Males (n = 36);\nNeutral Females (n = 46), Neutral Males (n = 47);\nHappy Females (n = 223), Happy Males (n = 203)",  
 x = "Gender",  
 y = "Percentage of Faithfulness") +  
 theme(plot.title = element\_text(face = "bold")) +  
   
 # The next few lines uses a theme dark template and overrides with grid design and legend outline  
   
 theme\_dark() +  
 theme(legend.box.background = element\_rect(colour = "black")) +  
 theme(  
 panel.background = element\_rect(fill = "#D7CADEFF",  
 colour = "black",  
 size = 0.5, linetype = 1),  
 panel.grid.major = element\_line(size = 0.5, linetype = 'solid',  
 colour = "black"),  
 panel.grid.minor = element\_line(size = 0.5, linetype = 'dotted',  
 colour = "black")  
 ) +  
   
 # The geom\_text function adds the value of percentages to each proportion of the bars; from "%0.2f%%", we print it with two decimals precision, calculating it using the count for each category of faithfulness, divided by the vector of total unhappy, neutral, and happy males and females  
   
 geom\_label(aes(label=sprintf("%0.2f%%", 100\*stat(count)/totalsUnhappy)), stat='count', position='fill', size = 3.2, show.legend = FALSE, fontface = "bold") +  
   
 # the next line turns the y-axis into a percentage instead of a ratio  
   
 scale\_y\_continuous(labels = scales::percent) +  
   
 # the next line makes the split in the graphs between unhappy, neutral, and happy females and males  
   
 facet\_grid(~rating)

