Data-Visualization-Assignment

Jeswin Paul & Yun Kai Lin

Group L

2020-11-04

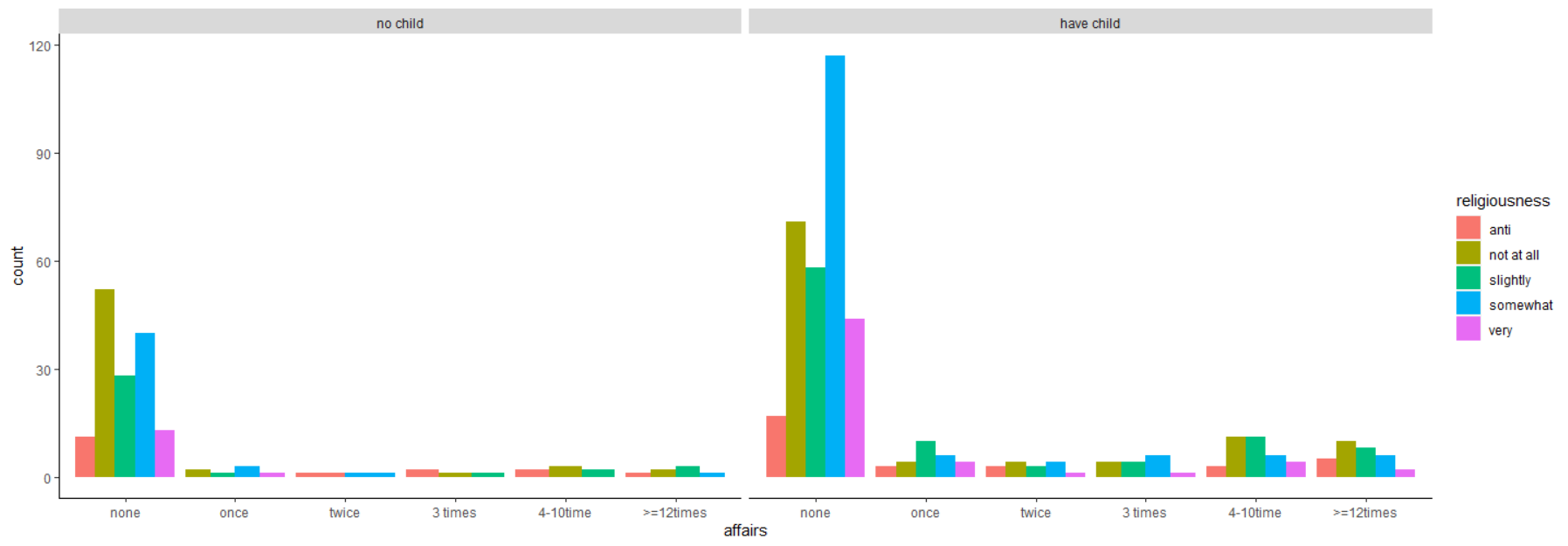
**Description of the data:**

This data represents data on married people, some characteristics of themselves and their marriage, and how many times they have been involved in extramarital affairs. This data is from a survey conducted by Psychology Today in 1969 regarding infidelity among married couples.

**Questions for visualization:**

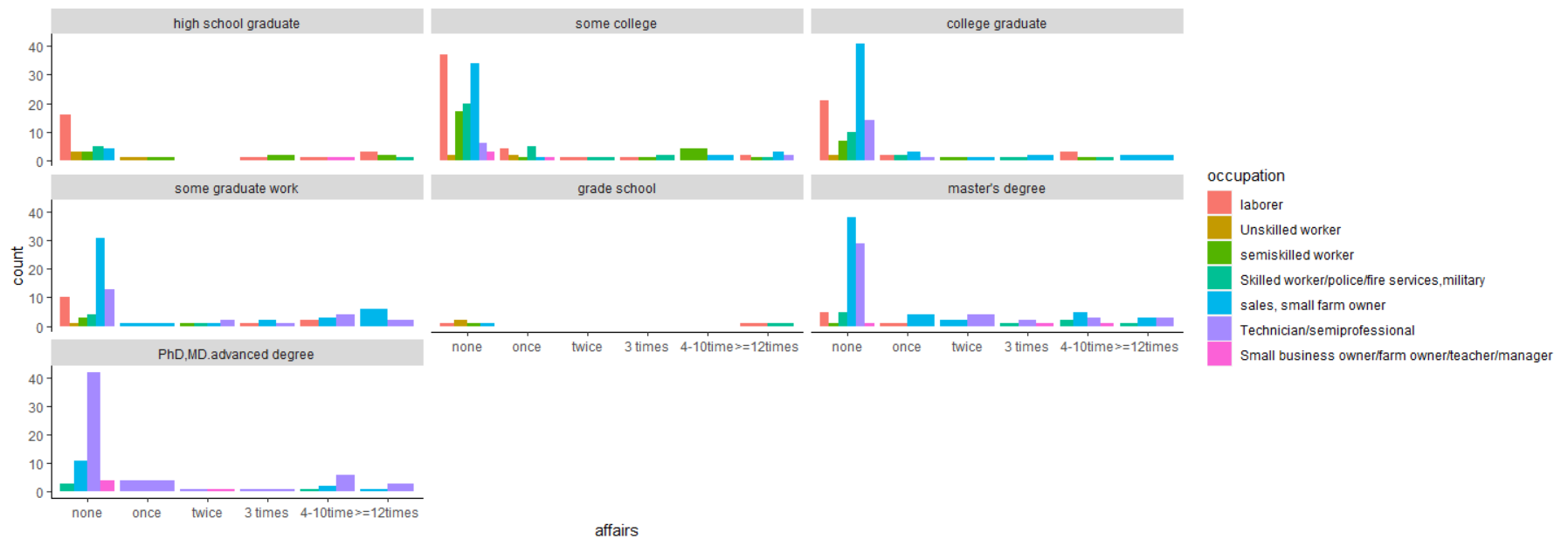
Q1. How does religiousness in a married person’s life correlate with how many extramarital affairs they have, is it possible that children in the relationship are tied with these two variables.

The reason that we chose to compare both religiousness and the presence of children to the number of affairs that a person has is because of the weight that those factors have when thinking of having an affair. The idea of extramarital affairs is frowned upon by religions as the act of marriage is considered a sacred act and having an affair violates the sacred nature of that union. This is not to say that without religion, society would tolerate it, but to say that religions have put weight on the union of marriage and the consequences of violating it. It goes without saying that having children in a relationship is a huge responsibility and can pressure people into keeping a healthy relationship, but that pressure can also cause a tear in relationships. Religions value the family unit heavily and particularly how a parent treats their children, extramarital affairs compromises both the marriage and the family unit. Do religious people with children tend to have less affairs than a non-religious person with no children? Are the different natures of the obligations a factor in whether a person chooses to have affairs?



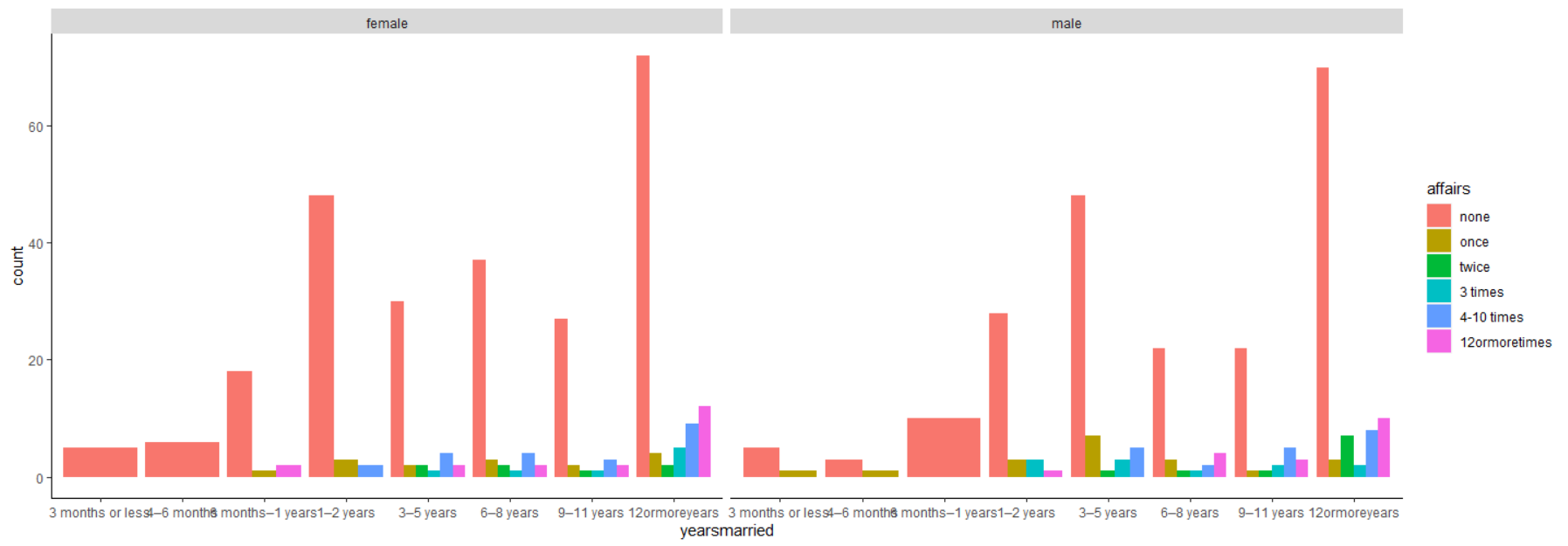
Q2. Does the level of education and occupation of a married person predict how many extramarital affairs they have?

A person’s education and occupation are good signs of socioeconomic status (SES), although a great education does not always equal to a higher-level occupation. It would be interesting to see how they are tied to the number of extramarital intercourses a person has. Is it that people who are less educated are more likely to cheat on their partners than people who is highly educated? This is not to say that we think a person’s intelligence is a determinate of whether they cheat or not, but it is to ask if the level of schooling is tied with number of affairs. Levels of occupation could be related to standings in social hierarchies, maybe if a person is in a prestigious position, they may consider cheating more than a person who is in a lower occupation level?



Q3. As couples are in a relationship for longer, do men tend to cheat more than women?

We wanted to see if there is a gender difference in how much extramarital intercourse a person had depending on how long the relationship has gone for. Does one gender tend to cheat more at the beginning of a relationship? Later into marriages, is one gender more likely to have an affair than the other?



R code:

library(psych) # load psych package for

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

# descriptive stats  
library(tidyverse) # load packages for importing,

## -- 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 ggplot2::%+%() masks psych::%+%()  
## x ggplot2::alpha() masks psych::alpha()  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

# cleaning, and plotting data  
library(here) # to load data

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

## here() starts at C:/Users/Thomas/Desktop

#import data  
Affairs <- read\_csv(file = here("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()  
## )

#mutate code to corresponded descriptions  
Affairs <- mutate(Affairs,  
 affairs = as.character(affairs),  
 affairs = fct\_recode(affairs,  
 "none" = "0",  
 "once" = "1",  
 "twice" = "2",  
 "3 times" = "3",  
 "4-10time" = "7",  
 ">=12times"= "12"))  
Affairs <- mutate(Affairs,  
 yearsmarried = as.character(yearsmarried),  
 yearsmarried = fct\_recode(yearsmarried,  
 "3 months or less" = "0.125",  
 "4–6 months" = "0.417",  
 "6 months–1 years" = "0.75",  
 "1–2 years" = "1.5",  
 "3–5 years" = "4",  
 "6–8 years" = "7",  
 "9–11 years"= "10",  
 "12ormoreyears"="15"))  
Affairs <- mutate(Affairs,  
 religiousness = as.character(religiousness),  
 religiousness = fct\_recode(religiousness,  
 "anti" = "1",  
 "not at all" = "2",  
 "slightly" = "3",  
 "somewhat" = "4",  
 "very" = "5"))  
Affairs <- mutate(Affairs,  
 education = as.character(education),  
 education = fct\_recode(education,  
 "grade school" = "9",  
 "high school graduate" = "12",  
 "some college" = "14",  
 "college graduate" = "16",  
 "some graduate work" = "17",  
 "master's degree" = "18",  
 "PhD,MD.advanced degree" = "20"))  
Affairs <- mutate(Affairs,  
 children = as.character(children),  
 children = fct\_recode(children,  
 "no child" = "no",  
 "have child" = "yes"  
 ))  
Affairs <- mutate(Affairs,  
 occupation = as.character(occupation),  
 occupation = fct\_recode(occupation,  
 "laborer" = "1",  
 "Unskilled worker" = "2",  
 "semiskilled worker" = "3",  
 "Skilled worker/police/fire services,military" = "4",  
 "sales, small farm owner" = "5",  
 "Technician/semiprofessional" = "6",  
 "Small business owner/farm owner/teacher/manager"= "7"  
 ))  
  
# select variables in use  
Affairs\_variables\_inuse <- select(Affairs,  
 affairs, yearsmarried,   
 religiousness, education, children,  
 gender, occupation)  
  
# select() all of the categorical data, and summarize  
Affairs\_variables\_inuse %>%   
 select(education, religiousness,  
 yearsmarried, affairs, children, gender, occupation) %>%   
 summary()

## education religiousness yearsmarried  
## high school graduate : 44 anti : 48 12ormoreyears :204   
## some college :154 not at all:164 3–5 years :105   
## college graduate :115 slightly :129 1–2 years : 88   
## some graduate work : 89 somewhat :190 6–8 years : 82   
## master's degree :112 very : 70 9–11 years : 70   
## PhD,MD.advanced degree: 80 6 months–1 years: 31   
## grade school : 7 (Other) : 21   
## affairs children gender   
## none :451 no child :171 Length:601   
## once : 34 have child:430 Class :character   
## 12ormoretimes: 38 Mode :character   
## twice : 17   
## 3 times : 19   
## 4-10 times : 42   
##   
## occupation   
## laborer :113   
## Unskilled worker : 13   
## semiskilled worker : 47   
## Skilled worker/police/fire services,military : 68   
## sales, small farm owner :204   
## Technician/semiprofessional :143   
## Small business owner/farm owner/teacher/manager: 13

# Sorting variables by factor ordering  
Affairs\_variables\_inuse$affairs <- factor(Affairs\_variables\_inuse$affairs,  
 levels = c("none", "once", "twice", "3 times",   
 "4-10time", ">=12times"))  
Affairs\_variables\_inuse$education <- factor(Affairs\_variables\_inuse$education,  
 levels = c("high school graduate", "some college",  
 "college graduate", "some graduate work",   
 "grade school", "master's degree",   
 "PhD,MD.advanced degree"))  
Affairs\_variables\_inuse$religiousness <- factor(Affairs\_variables\_inuse$religiousness,  
 levels = c("anti", "not at all", "slightly", "somewhat",   
 "very"))  
Affairs\_variables\_inuse$yearsmarried <- factor(Affairs\_variables\_inuse$yearsmarried,  
 levels = c("3 months or less", "4–6 months",  
 "6 months–1 years", "1–2 years",   
 "3–5 years", "6–8 years", "9–11 years",  
 "12ormoreyears"))  
Affairs\_variables\_inuse$occupation <- factor(Affairs\_variables\_inuse$occupation,  
 levels = c("laborer", "Unskilled worker",  
 "semiskilled worker", "Skilled worker/police/fire services,military",   
 "sales, small farm owner", "Technician/semiprofessional",   
 "Small business owner/farm owner/teacher/manager"  
 ))  
  
  
  
# create bar graph to show relationship between years of marriage, gender and affairs  
t2<-ggplot(data = Affairs\_variables\_inuse,  
 mapping = aes(x = yearsmarried, fill = affairs)) +  
 geom\_bar(position = position\_dodge()  
 )+  
 facet\_wrap(~gender)  
t2 + theme(  
 # Hide panel borders and remove grid lines  
 panel.border = element\_blank(),  
 panel.grid.major = element\_blank(),  
 panel.grid.minor = element\_blank(),  
 panel.background = element\_blank(),  
 # Change axis line  
 axis.line = element\_line(colour = "black")  
)

#create bar graph to show relationships between education, affairs...  
#and occupation  
t1 <-ggplot(Affairs\_variables\_inuse, aes(x =affairs ))+  
 geom\_bar(  
 aes(fill = occupation),  
 position = position\_dodge()  
 )+  
 facet\_wrap(~education)   
t1 + theme(  
 # Hide panel borders and remove grid lines  
 panel.border = element\_blank(),  
 panel.grid.major = element\_blank(),  
 panel.grid.minor = element\_blank(),  
 panel.background = element\_blank(),  
 # Change axis line  
 axis.line = element\_line(colour = "black")  
)

#create bar graph to show relationships between religiousness, affairs...  
#and whether have children  
 t <- ggplot(Affairs\_variables\_inuse, aes(x = affairs))+  
 geom\_bar(  
 aes(fill = religiousness),  
 position = position\_dodge()  
 )+  
 facet\_wrap(~children)   
   
 t + theme(  
 # Hide panel borders and remove grid lines  
 panel.border = element\_blank(),  
 panel.grid.major = element\_blank(),  
 panel.grid.minor = element\_blank(),  
 panel.background = element\_blank(),  
 # Change axis line  
 axis.line = element\_line(colour = "black")  
 )

**Limitations**

The biggest limitation we had with the data that was that the majority of the variables were coded so that they were categorical and not continuous in cases where they could have been. Variables such as affairs, yearsmarried, and age (not used) could have been coded as continuous and allowed for better visualized graphs. This limitation stretched into how we visualized the data as we were limited in the forms of graphs we could make. We settled for bra graphs for each of our questions, if some of the variables were coded differently, we could have visualized our questions differently such as using either a histogram or boxplot.