Stat/CS 187: Final Project

Huaye Yang, Halina Vercessi, and Matthew Mccarthy

Spring 2022

Table of Contents

knitr::opts\_chunk$set(echo = TRUE,  
 fig.width=12,   
 fig.height=10)  
  
# Load the packages  
pacman::p\_load(gapminder, tidyverse, skimr, socviz, grid, stringr,  
 usmap, maps, statebins, viridis, leaflet, cowplot, gridExtra, rpart, rpart.plot, caret, caTools)

# 1. Introduction

Before this survey, the most detailed data and reports on this subject were gathered and made in 2004 by the National Sleep Foundation. The data from the National Sleep Foundation provided the numbers on how many couples slept separately but did not gather data that could offer insight as to the reasons couples sleep apart and why couples sleep in separate beds. This data is interesting because it is the kind of data that could be useful to psychologists and social workers. This data can be used to examine stereotypes we have about why couples sleep separately and which age ranges of couples sleep separately along with being useful for examining why couples sleep separately.The data comes from a survey created in 2014 by Mona Chalabi as research for an article for Five Thirty Eight (from ABC News). The population from which this sample was drawn includes 1,057 American adults who were married, in a domestic partnership, in civil union or cohabiting with a significant other. The sample was selected with the help of SurveyMonkey, which gathered all 1,057 responses.

The participants only need to check boxes that they think best describes their situation. Bias can easily occur in this survey because:

1. It may be hard for participants to find a certain categorical box within each question to best describe their situation. In this circumstances, samples that are considered as outlier will fall into one of the categorical values. Also, we are only able to do a rough analysis based on designed categorical values. It can’t provide detailed information or quantitative analysis for accurate numerical values.
2. Participants may overly/underly describe their situation due to the form of survey. What we have in dataset is subjective, not objective.
3. We are not sure whether this survey is voluntary. If this is forced or hard for participants to refuse, the data can be messed because participants may randomly select the boxes instead of actually looking into the questions and participate. Also, we can’t guarantee how reliable the data is. From the recorded time (EndDate - StartDate),participants only spent around 3 minutes for answering 29 questions (Some of them only use less than 1 minute). There might be part of the data that are not meaningful. If this survey is voluntary (or maybe taken at a counselling institution,etc), participants who are willing to answer these questions may already have issues or at least went through the situation that were asked in the survey. This means we are not getting randomly distributed samples.

The cleaning that was done on the data was to rename the columns because the original column names were built from the survey questions (separated by periods) and they were very long and tedious to work with. Additionally, during the building of the following graphs, responses that were left blank were omitted to focus on relevant responses. We also removed the first row as it served as almost a second title row but was being counted in the rows of data. We considered labeling the blank / no answer spaces as N/A but N/A means something in this data as it is a valid response to specific check boxes.

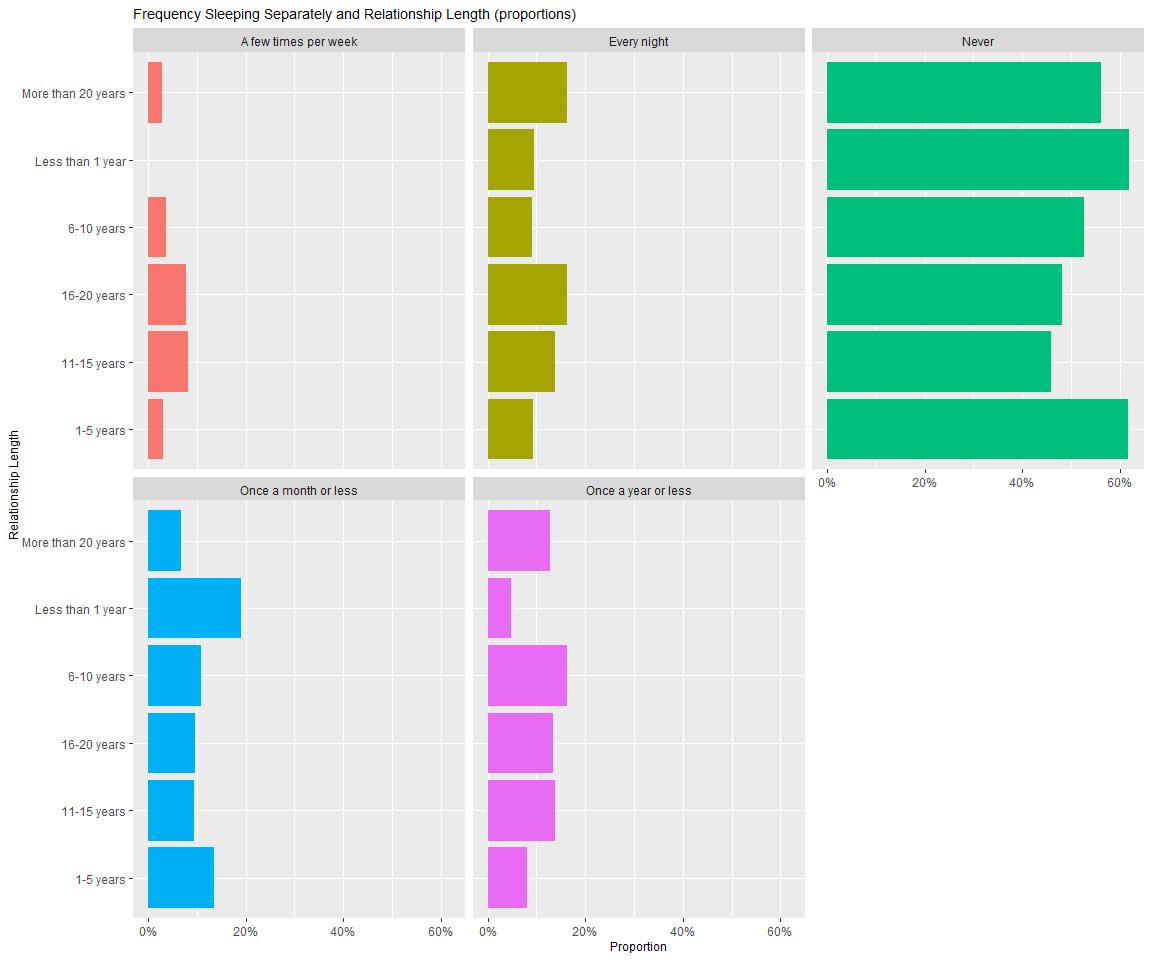
# Read in data  
sleeping\_alone\_data1 <- read.csv("R\_sleeping\_alone.csv", stringsAsFactors = T)  
  
# Rename columns to simpler names based on survey questions  
sleeping\_alone\_data <- sleeping\_alone\_data1 %>%  
   
 rename( Relationship\_Status = Which.of.the.following.best.describes.your.current.relationship.status.,   
 Relationship\_Length = How.long.have.you.been.in.your.current.relationship..If.you.are.not.currently.in.a.relationship..please.answer.according.to.your.last.relationship.,  
 How\_Often\_Sep\_Beds = When.both.you.and.your.partner.are.at.home..how.often.do.you.sleep.in.separate.beds.,   
 Diff\_Bed\_Where\_You\_Sleep = When.you.re.not.sleeping.in.the.same.bed.as.your.partner..where.do.you.typically.sleep.,   
 Diff\_Bed\_Where\_Partner\_Sleep = When.you.re.not.sleeping.in.the.same.bed..where.does.your.partner.typically.sleep.,   
 Sep\_Beds\_Reasons = What.are.the.reasons.that.you.sleep.in.separate.beds..Please.select.all.that.apply.,   
 When\_First\_Time\_Sep\_Beds = When.was.the.first.time.you.slept.in.separate.beds.,   
 Statement\_Help\_Stay\_Together = To.what.extent.do.you.agree.with.the.following.statement...sleeping.in.separate.beds.helps.us.to.stay.together..,   
 Statement\_Better\_Sleep = To.what.extent.do.you.agree.with.the.following.statement...we.sleep.better.when.we.sleep.in.separate.beds..,  
 Statement\_Improved\_Sex\_Life = To.what.extent.do.you.agree.with.the.following.statement.ë\_.our.sex.life.has.improved.as.a.result.of.sleeping.in.separate.beds..ë\_,   
 Current\_Occupation = Which.of.the.following.best.describes.your.current.occupation.,  
 Household\_Income = Household.Income,   
 Location\_Census\_Region = Location..Census.Region.)  
   
  
sleeping\_alone\_data <- sleeping\_alone\_data %>% slice(-c(1))

# 2. Data Visualizations

## 2.1 Length of Relationship and Age:

### 2.1.1 Graph 1

# Prepare data  
sleeping\_alone\_data2 <- sleeping\_alone\_data %>%  
   
 filter(Relationship\_Length != "", How\_Often\_Sep\_Beds != "") %>%  
   
 group\_by(Relationship\_Length, How\_Often\_Sep\_Beds) %>%  
   
 summarise(count\_freq = n()) %>%  
   
 mutate(Proportion = round(count\_freq/sum(count\_freq), digits = 5)) %>%  
   
 select(Relationship\_Length, How\_Often\_Sep\_Beds, Proportion) %>%   
   
 slice(-c(1)) %>%  
   
 ungroup()  
  
# Bar graph 1: relationship length and frequency sleeping separately   
rel\_length\_bar <- ggplot(data = sleeping\_alone\_data2,  
 mapping = aes(x = Relationship\_Length,  
 y = Proportion,   
 fill = How\_Often\_Sep\_Beds)) +  
   
 geom\_col(position = "dodge") +   
   
 labs(x = "Relationship Length",  
 fill = "How Often Couples Sleep Separately",   
 title = "Frequency Sleeping Separately and Relationship Length (proportions)")+  
   
 scale\_y\_continuous(labels = scales::percent) +  
 facet\_wrap(~How\_Often\_Sep\_Beds) +  
 theme(legend.position = "NA",   
 title = element\_text(size = 9)) +  
 coord\_flip()  
  
rel\_length\_bar

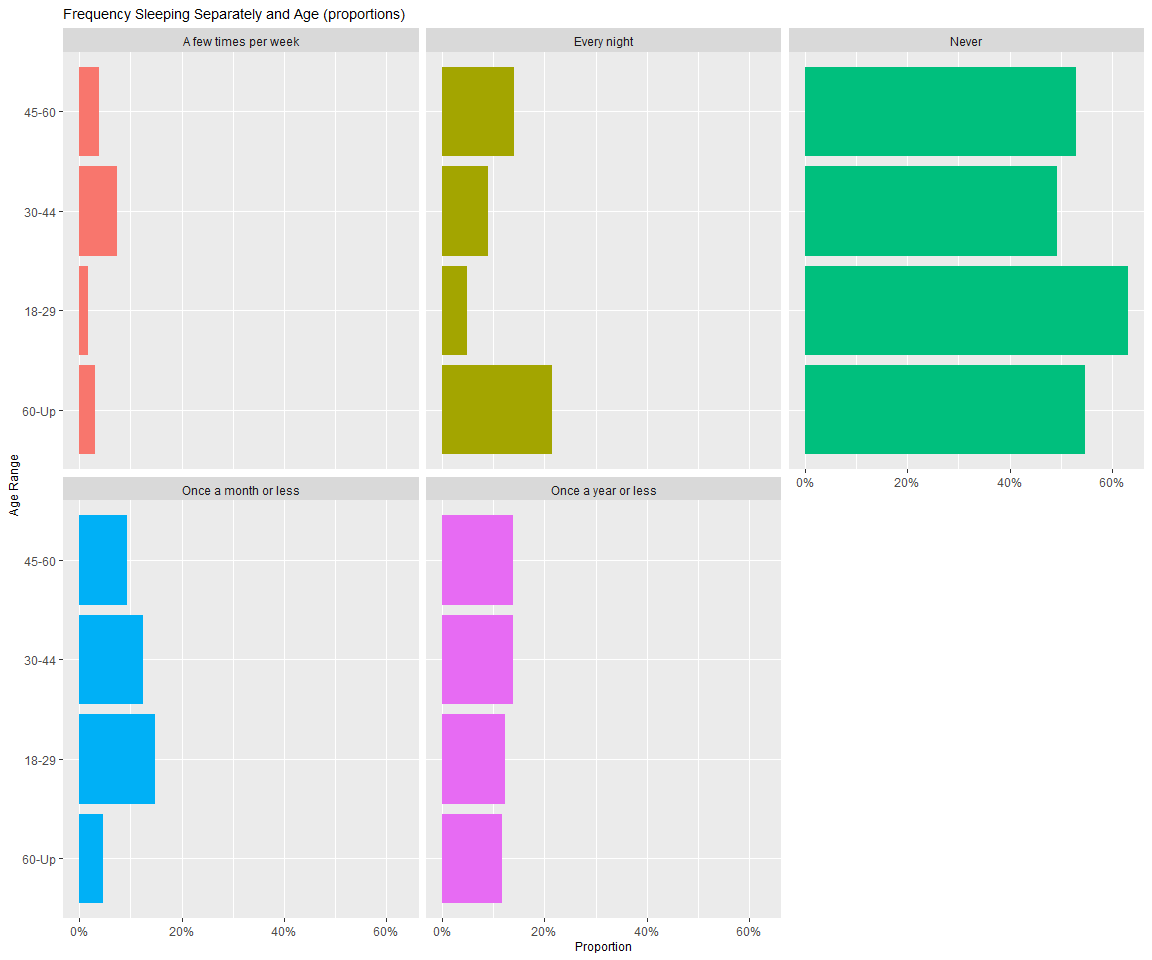


### 2.1.2 Graph 1.1

# Prepare data  
sleeping\_alone\_data3 <- sleeping\_alone\_data %>%  
   
 filter(Age != "", How\_Often\_Sep\_Beds != "") %>%  
   
 group\_by(Age, How\_Often\_Sep\_Beds) %>%  
   
 summarise(count\_freq = n()) %>%  
   
 mutate(Proportion = round(count\_freq/sum(count\_freq), digits = 5)) %>%  
   
 select(Age, How\_Often\_Sep\_Beds, Proportion) %>%   
   
 slice(-c(1)) %>%  
   
 ungroup()

## `summarise()` has grouped output by 'Age'. You can override using the `.groups` argument.

# Bar graph 2: Age and frequency sleeping separately  
prop\_age\_bar <- ggplot(data = sleeping\_alone\_data3,  
 mapping = aes(x = Age,  
 y = Proportion,   
 fill = How\_Often\_Sep\_Beds)) +  
   
 geom\_col(position = "dodge") +   
   
 labs(x = "Age",  
 fill = "How Often Couples Sleep Separately",   
 title = "Frequency Sleeping Separately and Age (proportions)")+  
   
 scale\_y\_continuous(labels = scales::percent)+   
 scale\_x\_discrete(name = "Age Range", labels = c("60-Up", "18-29", "30-44", "45-60")) +  
 facet\_wrap(~How\_Often\_Sep\_Beds) +  
 theme(legend.position = "NA",   
 title = element\_text(size = 9)) +  
 coord\_flip()  
  
prop\_age\_bar



### 2.1.3 Summarizing Paragraph:

From these two graphs, it is clear that the majority of couples who took the survey do not sleep separately. From the first graph, around 18% of couples who have been in a relationship for 16-20 years and more than 20 years do sleep in separate beds. These percentages are on the higher end for proportions of couples that have slept or do sleep separately. This aspect of the data fits our society’s stereotyping that couples who have been together longer are more likely to sleep separately. Another notable percentage is that almost 20% of couples who have been together for less than a year reported sleeping separately once a month or less.

These graphs communicate and fit what the public might expect to see from a survey like this, however, it does not give any indication as to the reasons couples who have been together longer might be sleeping separately. The following sections and figures will aim to further examine this aspect of the population sample. More specifically, could occupation and income (which are often tied to how busy or occupied a person might be in their daily activities) be related to couple sleeping separately? Additionally, what are the reasons for certain categories of couples (older, younger, etc) sleeping separately and have they noticed changes in their relationship from sleeping separately?

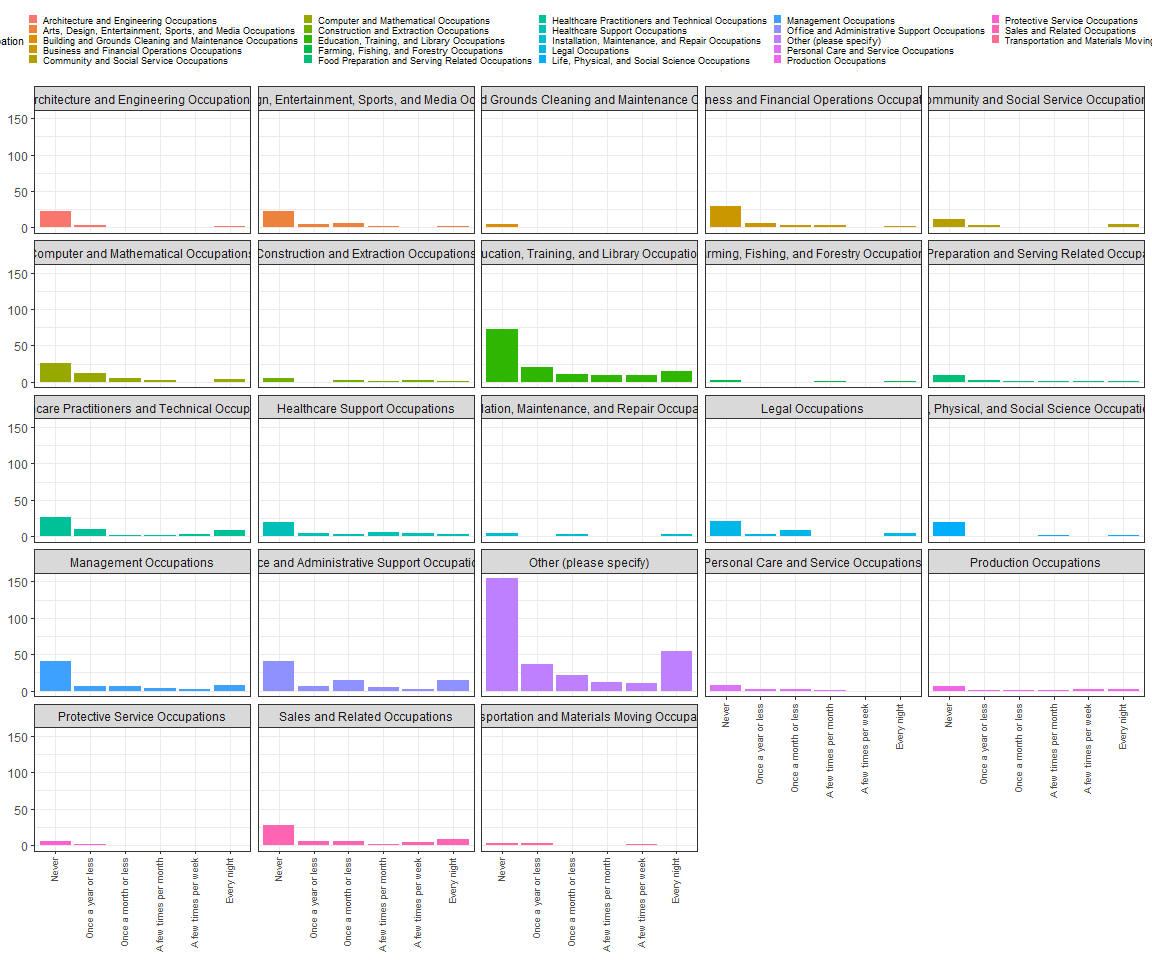
## 2.2 Occupation and Income

### 2.2.1 Graph 2

# change the default theme below:  
theme\_set(theme\_bw())  
# take a look at the categories of How\_Often\_Sep\_Beds  
summary(sleeping\_alone\_data$How\_Often\_Sep\_Beds)

## A few times per month A few times per week   
## 14 62 46   
## Every night Never Once a month or less   
## 147 586 101   
## Once a year or less Response   
## 137 0

# change How\_Often\_Sep\_Beds into an ordered factor  
rank<-as.factor(sleeping\_alone\_data$How\_Often\_Sep\_Beds)  
sleeping\_alone\_data4 <- sleeping\_alone\_data %>%  
 mutate(How\_Often\_Sep\_Beds = factor(rank, ordered = TRUE,   
 levels = c("Never",  
 "Once a year or less",  
 "Once a month or less",  
 "A few times per month",  
 "A few times per week",  
 "Every night")))  
# Relate occupation to how often they sleep separately  
sleeping\_alone\_data4 %>%  
 filter(Current\_Occupation != "") %>%   
 ggplot(aes(x = How\_Often\_Sep\_Beds,  
 fill = Current\_Occupation)) +  
 geom\_bar()+  
   
 guides(fill=guide\_legend(title="Occupation"))+  
 facet\_wrap(~Current\_Occupation)+   
   
 theme(axis.text.x = element\_text(angle = 90,  
 hjust = 1,  
 vjust = 0.25,  
 size = 7),  
 legend.position = "top",  
 legend.text = element\_text(size = 7),  
 legend.key.size = unit(0.25,'cm'),  
 legend.title = element\_text(size= 8),  
 axis.title.x = element\_blank(),  
 axis.title.y = element\_blank())



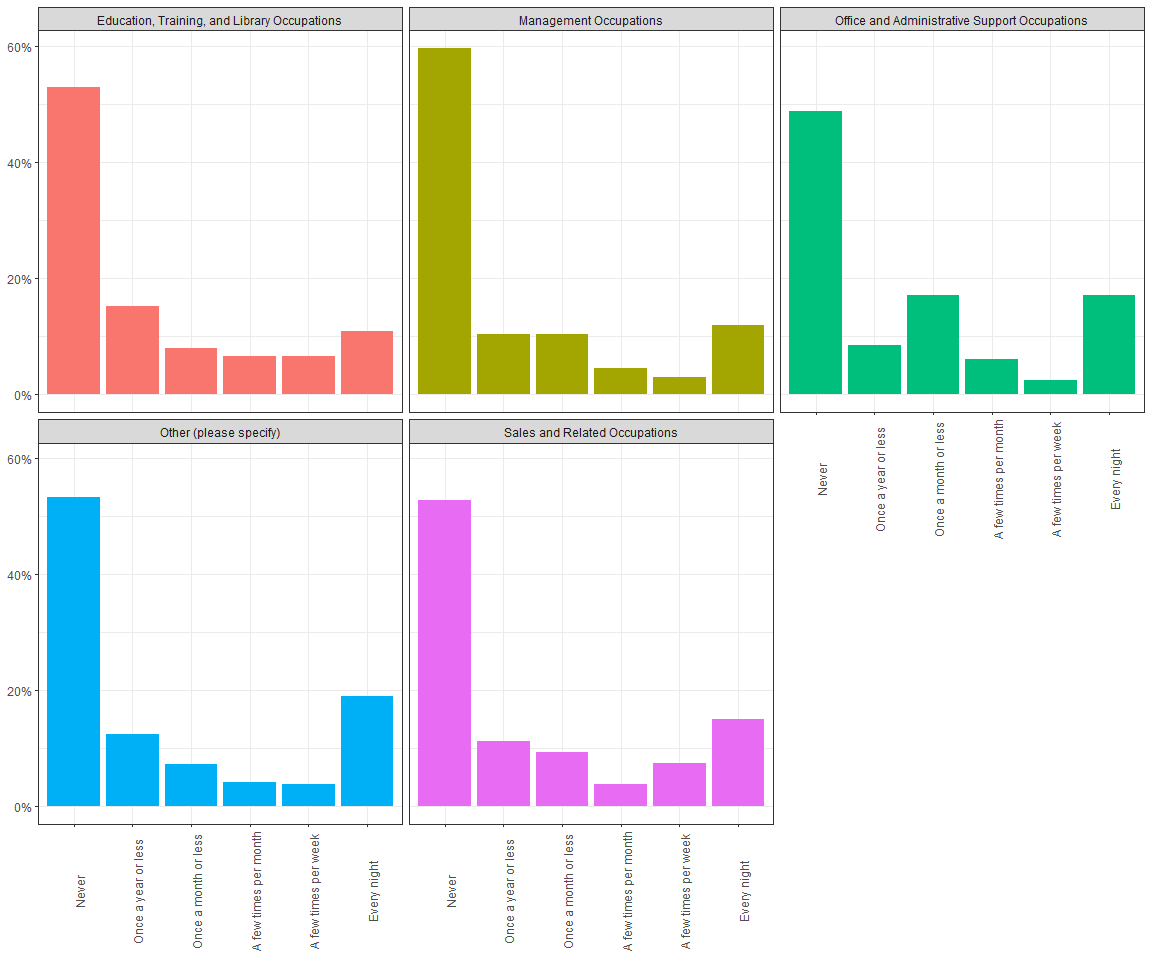
# summary() to see each category name and ready to draw plots for top 5   
# of the categories.   
summary(sleeping\_alone\_data$Current\_Occupation)

##   
## 29   
## Architecture and Engineering Occupations   
## 30   
## Arts, Design, Entertainment, Sports, and Media Occupations   
## 39   
## Building and Grounds Cleaning and Maintenance Occupations   
## 7   
## Business and Financial Operations Occupations   
## 46   
## Community and Social Service Occupations   
## 22   
## Computer and Mathematical Occupations   
## 51   
## Construction and Extraction Occupations   
## 13   
## Education, Training, and Library Occupations   
## 138   
## Farming, Fishing, and Forestry Occupations   
## 5   
## Food Preparation and Serving Related Occupations   
## 16   
## Healthcare Practitioners and Technical Occupations   
## 51   
## Healthcare Support Occupations   
## 39   
## Installation, Maintenance, and Repair Occupations   
## 12   
## Legal Occupations   
## 37   
## Life, Physical, and Social Science Occupations   
## 25   
## Management Occupations   
## 67   
## Office and Administrative Support Occupations   
## 82   
## Other (please specify)   
## 289   
## Personal Care and Service Occupations   
## 13   
## Production Occupations   
## 14   
## Protective Service Occupations   
## 7   
## Response   
## 0   
## Sales and Related Occupations   
## 53   
## Transportation and Materials Moving Occupations   
## 8

# create a new dataset which contains the top 5(ranked from number  
# of participants) occupation  
sleeping\_alone\_data4\_1 <- sleeping\_alone\_data4 %>%  
 select(Current\_Occupation,How\_Often\_Sep\_Beds) %>%  
 filter(Current\_Occupation == "Education, Training, and Library Occupations"|  
 Current\_Occupation == "Other (please specify)"|  
 Current\_Occupation == "Office and Administrative Support Occupations"|  
 Current\_Occupation == "Management Occupations"|  
 Current\_Occupation == "Sales and Related Occupations")   
# create a proportion variable for plotting  
sleeping\_alone\_data4\_1 <- sleeping\_alone\_data4\_1 %>%  
   
 group\_by(Current\_Occupation, How\_Often\_Sep\_Beds) %>%   
   
 summarise(freq = n()) %>%   
   
 mutate(Prop = round(freq/sum(freq), digits = 5)) %>%   
   
 select(Current\_Occupation, How\_Often\_Sep\_Beds, Prop) %>%   
   
 ungroup()

## `summarise()` has grouped output by 'Current\_Occupation'. You can override using the `.groups` argument.

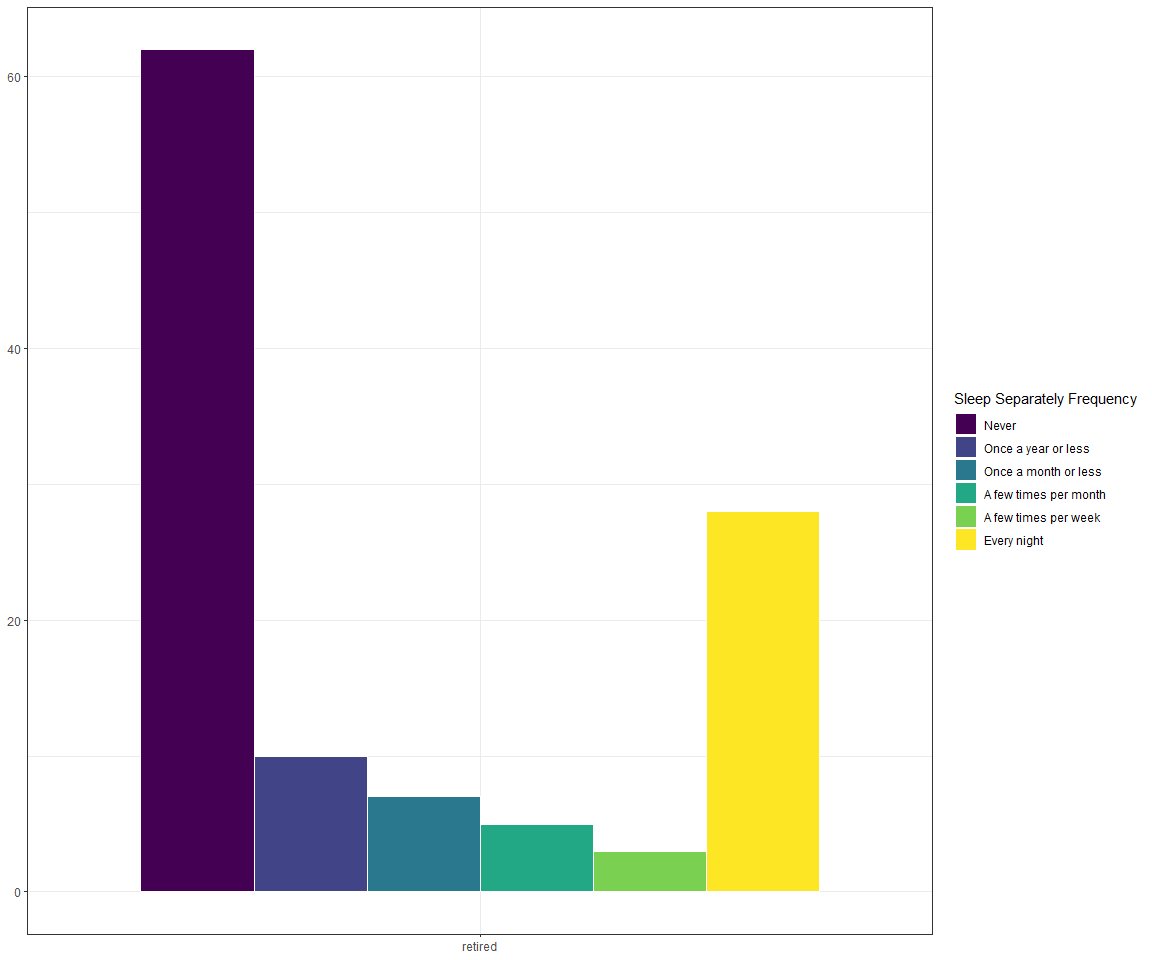
# make a bar plot and separate for top 5 occupations  
ggplot(data = sleeping\_alone\_data4\_1, aes(y = Prop,  
 x = How\_Often\_Sep\_Beds,  
 fill = Current\_Occupation))+  
   
 geom\_col(show.legend = F)+  
   
 scale\_y\_continuous(labels = scales::percent)+  
   
 facet\_wrap(~Current\_Occupation)+  
   
   
 theme(axis.text.x = element\_text(angle = 90),  
 axis.title.x = element\_blank(),  
 axis.title.y = element\_blank())



## for people who have further response instead of checking boxes  
## (choose to plot those who answered "retired" or "Retired)  
sleeping\_alone\_data4 %>%  
 filter(X.12 == "retired" | X.12 == "Retired")%>%  
 mutate(X.12 ="retired") %>%  
   
 ggplot(aes(x = X.12, fill= How\_Often\_Sep\_Beds)) +  
   
 geom\_bar(color="white",position = "dodge")+  
   
 guides(fill=guide\_legend(title="Sleep Separately Frequency"))+  
   
 theme(axis.text.x = element\_text("Retired"),  
 axis.title.x = element\_blank(),  
 axis.title.y = element\_blank())

## Warning in grid.Call(C\_stringMetric, as.graphicsAnnot(x$label)): font family not  
## found in Windows font database

## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font  
## family not found in Windows font database  
  
## Warning in grid.Call(C\_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font  
## family not found in Windows font database



### 2.2.2 Summarizing Paragraph

The first thing we need to notice from the first overall plot is the proportion of each occupation that our participants came from. Since they all have the same scale on y-axis, it is not hard to find the top 5 occupations(ranked in number of participants).Also, this gave us a general sense of the make-up for our dataset and how representative it is.We would like to look at the top 5 occupations in detail and give a more specific interpretation. The top 5 occupations, as we see in second plot, are: “Education, Training, and Library Occupations” (138 participants), “Other (please specify)” (289 participants), “Office and Administrative Support Occupations” (82 participants), “Management Occupations” (67 participants), “Sales and Related Occupations” (53 participants).

People who are in Education, Training, and Library Occupations seems to have the most response of “Never” sleeping separately(around 50% ~ 60%). Regardless of the occupation,the number of people who sleep separately is bigger than other frequencies.This might refer to a bias where participants may over/under estimate their circumstances while doing self-evaluation.

For people who checked the box of “Other(please specify)”, we select retired samples to draw a bar graph. Distribution of people who retired have heavy tails on both side(Never & Every night), but the number of people who never sleeping separately is almost twice as large as those who sleeping separately every night.

### 2.2.3 Graph 3

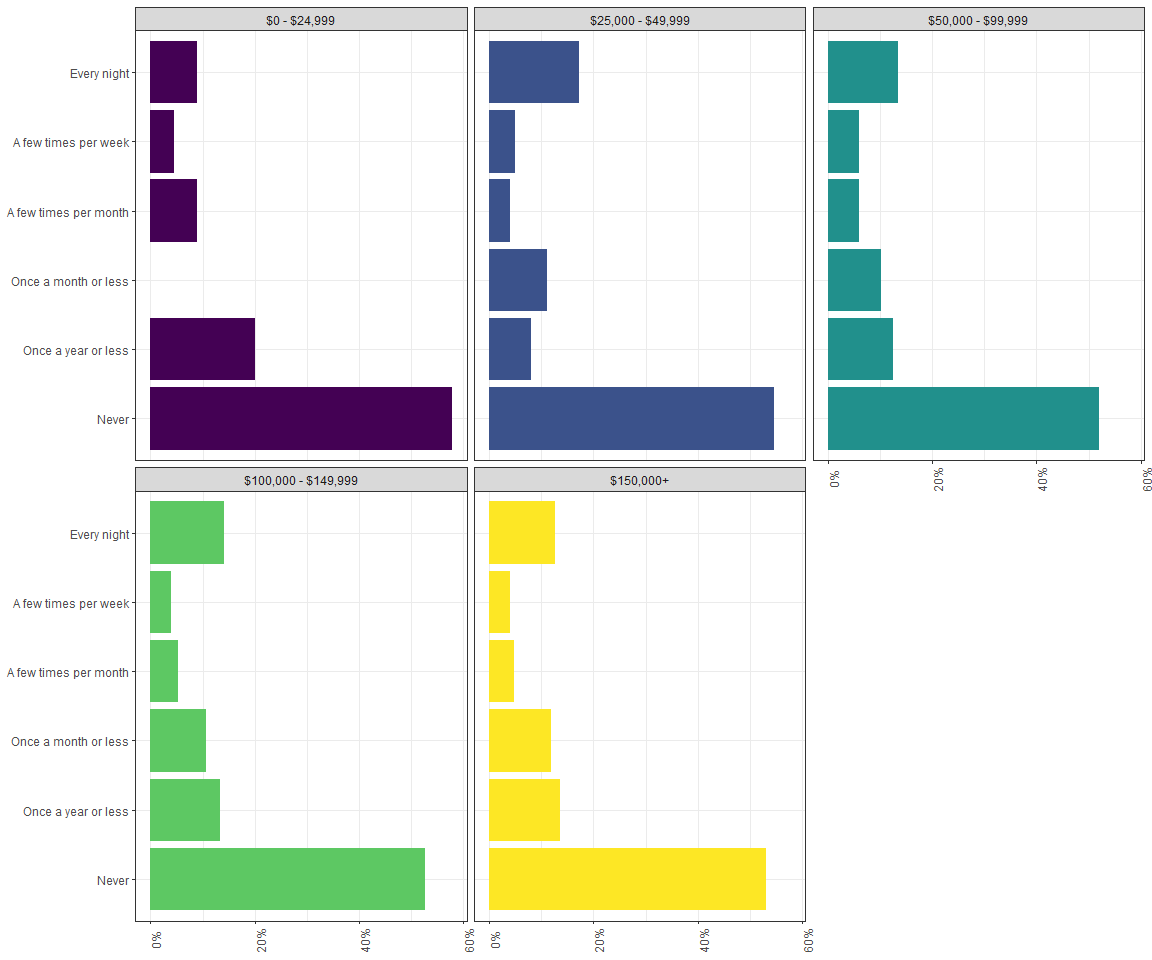
# take a look at the categories of Household\_Income  
summary(sleeping\_alone\_data4$Household\_Income)

## $0 - $24,999 $100,000 - $149,999 $150,000+   
## 283 45 226 126   
## $25,000 - $49,999 $50,000 - $99,999 Response   
## 99 314 0

# change Household\_Income into an ordered factor(in the dataset where   
# How\_Often\_Sep\_Beds alreay function as an ordered factor)  
rank<-as.factor(sleeping\_alone\_data4$Household\_Income)  
sleeping\_alone\_data4\_2 <- sleeping\_alone\_data4 %>%  
 mutate(Household\_Income = factor(rank, ordered = TRUE,   
 levels = c("NA",  
 "$0 - $24,999",  
 "$25,000 - $49,999",  
 "$50,000 - $99,999",  
 "$100,000 - $149,999",  
 "$150,000+",  
 "Response")))  
# create a proportion variable for plotting  
sleeping\_alone\_data4\_2 <- sleeping\_alone\_data4\_2 %>%  
   
 group\_by(Household\_Income, How\_Often\_Sep\_Beds) %>%   
   
 summarise(freq = n()) %>%   
   
 mutate(Prop = round(freq/sum(freq), digits = 5)) %>%   
   
 select(Household\_Income, How\_Often\_Sep\_Beds, Prop) %>%   
   
 ungroup()

## `summarise()` has grouped output by 'Household\_Income'. You can override using the `.groups` argument.

# make a bar plot and separate by Household\_Income  
sleeping\_alone\_data4\_2 %>%   
 filter(Household\_Income != "NA") %>%  
   
 ggplot(aes(y = Prop,  
 x = How\_Often\_Sep\_Beds,  
 fill = Household\_Income)) +  
 geom\_col(show.legend = F)+  
   
 scale\_y\_continuous(labels = scales::percent)+  
   
 theme(axis.text.x = element\_text(angle = 90),  
 axis.title.x = element\_blank(),  
 axis.title.y = element\_blank())+  
 coord\_flip()+  
 facet\_wrap(~Household\_Income)



### 2.2.4 Summarizing Paragraph:

When we relate household income to how often participants sleep separately with their partner, we can see that there is not a significant difference between groups with different levels of income. The proportion of answers for both tails takes more than half of the total number of answers. The number of answer tend to decrease as the frequency of sleeping separately increases, but not for both tails.

## 2.3 Reasons for Sleeping Separately

### 2.3.1 Graph 4

# GRAPH 4  
  
#Horizontal Bar Chart for Reason 1  
sleeping\_alone\_data5 <- sleeping\_alone\_data %>%   
   
 group\_by(Sep\_Beds\_Reasons, How\_Often\_Sep\_Beds) %>%   
   
 summarise(freq = n()) %>%   
   
 mutate(Prop = round(freq/sum(freq), digits = 5)) %>%   
   
 select(Sep\_Beds\_Reasons, How\_Often\_Sep\_Beds, Prop) %>%   
   
 slice(-c(1)) %>%   
   
 ungroup()

## `summarise()` has grouped output by 'Sep\_Beds\_Reasons'. You can override using the `.groups` argument.

BarOftenVsReason1 <- ggplot(data = sleeping\_alone\_data5,  
 mapping = aes(y = Prop,  
 x = How\_Often\_Sep\_Beds,  
 fill = How\_Often\_Sep\_Beds)) +  
   
 geom\_bar(stat = "identity",  
 show.legend = F) +   
 coord\_flip() +   
 labs(title = "One of Us Snores",  
 x = "",  
 y = "") +  
 scale\_y\_continuous(labels = scales::percent) +  
   
 theme(plot.title = element\_text(size = 15, hjust = 0.5))  
#Horizontal Bar Chart for Reason 2  
sleeping\_alone\_data6 <- sleeping\_alone\_data %>%   
   
 group\_by(X.2, How\_Often\_Sep\_Beds) %>%   
   
 summarise(freq = n()) %>%   
   
 mutate(Prop = round(freq/sum(freq), digits = 5)) %>%   
   
 select(X.2, How\_Often\_Sep\_Beds, Prop) %>%   
   
 slice(-c(1)) %>%   
   
 ungroup()

## `summarise()` has grouped output by 'X.2'. You can override using the `.groups` argument.

BarOftenVsReason2 <- ggplot(data = sleeping\_alone\_data6,  
 mapping = aes(y = Prop,  
 x = How\_Often\_Sep\_Beds,  
 fill = How\_Often\_Sep\_Beds)) +  
   
 geom\_bar(stat = "identity",  
 show.legend = F) +   
 coord\_flip() +   
 labs(title = "Frequent Bathroom Trips",  
 x = "",  
 y = "") +  
   
 scale\_y\_continuous(labels = scales::percent) +  
   
 theme(plot.title = element\_text(size = 15, hjust = 0.5))  
#Horizontal Bar Chart for Reason 3  
sleeping\_alone\_data7 <- sleeping\_alone\_data %>%   
   
 group\_by(X.4, How\_Often\_Sep\_Beds) %>%   
   
 summarise(freq = n()) %>%   
   
 mutate(Prop = round(freq/sum(freq), digits = 5)) %>%   
   
 select(X.4, How\_Often\_Sep\_Beds, Prop) %>%   
   
 slice(-c(1)) %>%   
   
 ungroup()

## `summarise()` has grouped output by 'X.4'. You can override using the `.groups` argument.

BarOftenVsReason3 <- ggplot(data = sleeping\_alone\_data7,  
 mapping = aes(y = Prop,  
 x = How\_Often\_Sep\_Beds,  
 fill = How\_Often\_Sep\_Beds)) +  
   
 geom\_bar(stat = "identity",  
 show.legend = F) +   
 coord\_flip() +   
 labs(title = "No Longer Physically Intimate",  
 x = "",  
 y = "") +  
   
 scale\_y\_continuous(labels = scales::percent) +  
   
 theme(plot.title = element\_text(size = 15, hjust = 0.5))  
#Horizontal Bar Chart for Reason 4  
sleeping\_alone\_data8 <- sleeping\_alone\_data %>%   
   
 group\_by(X.5, How\_Often\_Sep\_Beds) %>%   
   
 summarise(freq = n()) %>%   
   
 mutate(Prop = round(freq/sum(freq), digits = 5)) %>%   
   
 select(X.5, How\_Often\_Sep\_Beds, Prop) %>%   
   
 slice(-c(1)) %>%   
   
 ungroup()

## `summarise()` has grouped output by 'X.5'. You can override using the `.groups` argument.

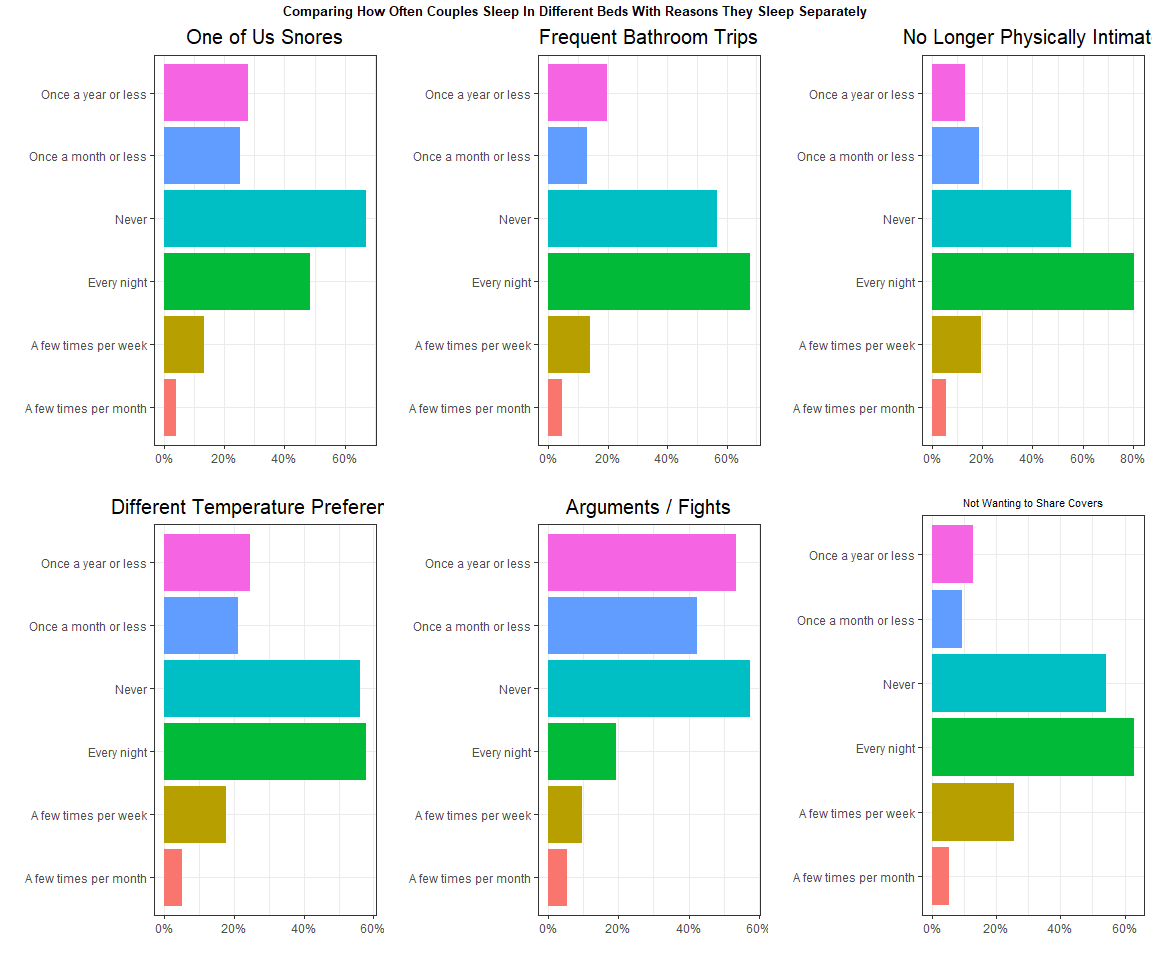
BarOftenVsReason4 <- ggplot(data = sleeping\_alone\_data8,  
 mapping = aes(y = Prop,  
 x = How\_Often\_Sep\_Beds,  
 fill = How\_Often\_Sep\_Beds)) +  
   
 geom\_bar(stat = "identity",  
 show.legend = F) +   
 coord\_flip() +   
 labs(title = "Different Temperature Preferences",  
 x = "",  
 y = "") +  
   
 scale\_y\_continuous(labels = scales::percent) +  
   
 theme(plot.title = element\_text(size = 15, hjust = 0.5))  
#Horizontal Bar Chart for Reason 5  
sleeping\_alone\_data9 <- sleeping\_alone\_data %>%   
   
 group\_by(X.6, How\_Often\_Sep\_Beds) %>%   
   
 summarise(freq = n()) %>%   
   
 mutate(Prop = round(freq/sum(freq), digits = 5)) %>%   
   
 select(X.6, How\_Often\_Sep\_Beds, Prop) %>%   
   
 slice(-c(1)) %>%   
   
 ungroup()

## `summarise()` has grouped output by 'X.6'. You can override using the `.groups` argument.

BarOftenVsReason5 <- ggplot(data = sleeping\_alone\_data9,  
 mapping = aes(y = Prop,  
 x = How\_Often\_Sep\_Beds,  
 fill = How\_Often\_Sep\_Beds)) +  
   
 geom\_bar(stat = "identity",  
 show.legend = F) +   
 coord\_flip() +   
 labs(title = "Arguments / Fights",  
 x = "",  
 y = "") +  
   
 scale\_y\_continuous(labels = scales::percent) +  
   
 theme(plot.title = element\_text(size = 15, hjust = 0.5))  
#Horizontal Bar Chart for Reason 6  
sleeping\_alone\_data10 <- sleeping\_alone\_data %>%   
   
 group\_by(X.8, How\_Often\_Sep\_Beds) %>%   
   
 summarise(freq = n()) %>%   
   
 mutate(Prop = round(freq/sum(freq), digits = 5)) %>%   
   
 select(X.8, How\_Often\_Sep\_Beds, Prop) %>%   
   
 slice(-c(1)) %>%   
   
 ungroup()

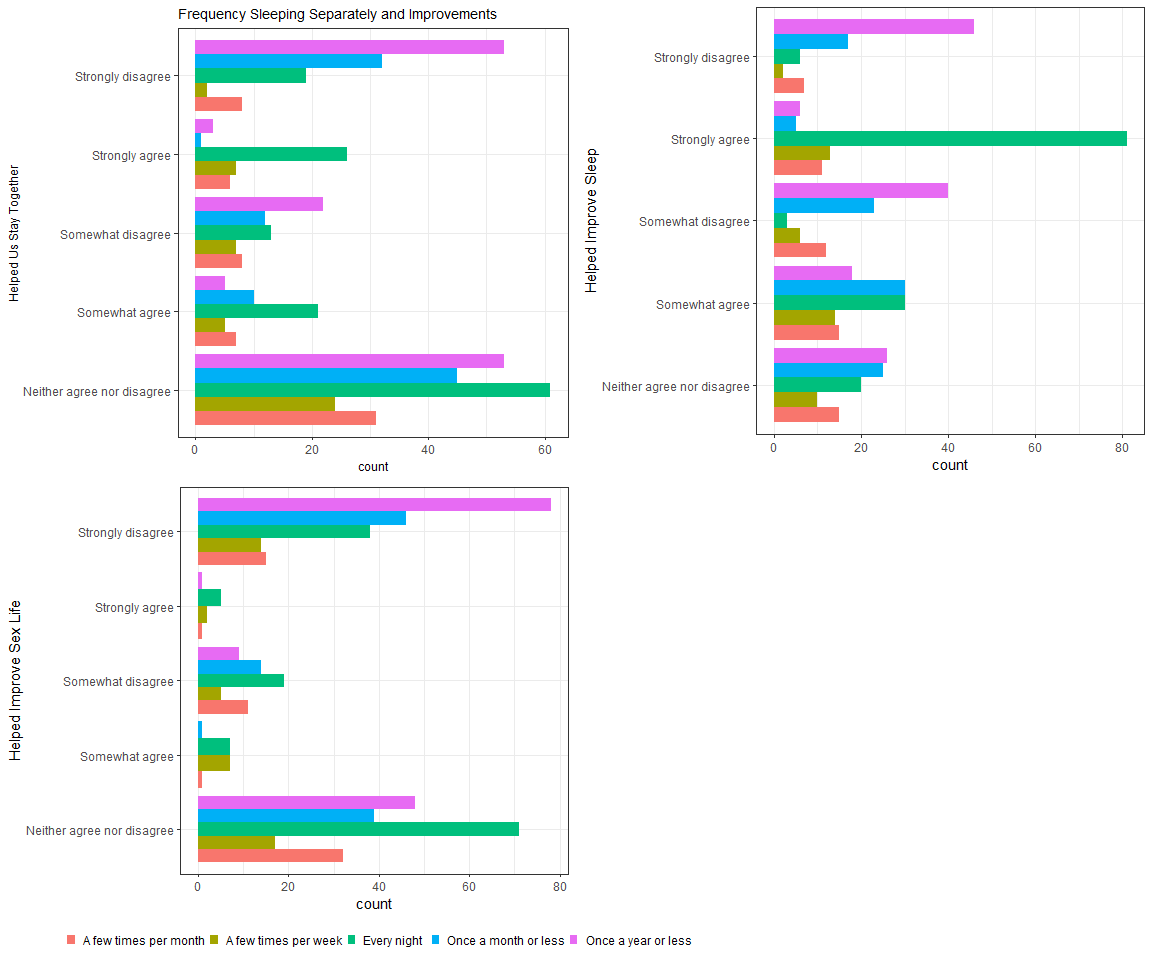
## `summarise()` has grouped output by 'X.8'. You can override using the `.groups` argument.

BarOftenVsReason6 <- ggplot(data = sleeping\_alone\_data10,  
 mapping = aes(y = Prop,  
 x = How\_Often\_Sep\_Beds,  
 fill = How\_Often\_Sep\_Beds)) +  
   
 geom\_bar(stat = "identity",  
 show.legend = F) +   
 coord\_flip() +   
 labs(title = "Not Wanting to Share Covers",  
 x = "",  
 y = "") +  
   
 scale\_y\_continuous(labels = scales::percent) +  
   
 theme(plot.title = element\_text(size = 8, hjust = 0.5))  
grid.arrange(BarOftenVsReason1, BarOftenVsReason2, BarOftenVsReason3, BarOftenVsReason4, BarOftenVsReason5, BarOftenVsReason6,nrow = 2,  
 top=textGrob("Comparing How Often Couples Sleep In Different Beds With Reasons They Sleep Separately",   
 gp=gpar (fontsize=10, fontface = "bold")))



### 2.3.2 Graph 4.1

help\_stay\_together <- sleeping\_alone\_data %>% filter(Statement\_Help\_Stay\_Together != "") %>%  
 ggplot(mapping = aes(x = Statement\_Help\_Stay\_Together,  
 fill = How\_Often\_Sep\_Beds)) +  
 geom\_bar(position = "dodge") +   
 labs(x = "Helped Us Stay Together",  
 fill = "How Often Couples Sleep Separately",  
 title = "Frequency Sleeping Separately and Improvements") +  
 theme(legend.position = "NA",   
 title = element\_text(size = 9)) +  
 coord\_flip()  
  
better\_sleep <- sleeping\_alone\_data %>% filter(Statement\_Better\_Sleep != "") %>%  
 ggplot(mapping = aes(x = Statement\_Better\_Sleep,  
 fill = How\_Often\_Sep\_Beds)) +  
 geom\_bar(position = "dodge") +   
 labs(x = "Helped Improve Sleep",  
 fill = "How Often Couples Sleep Separately") +  
 theme(legend.position = "NA" ) +  
 coord\_flip()  
  
improved\_sex\_life <- sleeping\_alone\_data %>% filter(Statement\_Improved\_Sex\_Life != "") %>%  
 ggplot(mapping = aes(x = Statement\_Improved\_Sex\_Life,  
 fill = How\_Often\_Sep\_Beds)) +  
 geom\_bar(position = "dodge") +   
 labs(x = "Helped Improve Sex Life",  
 fill = "How Often Couples Sleep Separately") +  
 theme(legend.position = "bottom",  
 legend.key.size = unit(0.25, 'cm'),   
 legend.title = element\_blank(),   
 legend.text = element\_text(size=9)) +  
 coord\_flip()  
  
  
  
plot\_grid(help\_stay\_together, better\_sleep, improved\_sex\_life, ncol = 2, nrow = 2, label\_size = 9)



### 2.3.3 Summarizing Paragraph:

These graphs compare how often couples sleep in different beds with some reasons why they sleep separately. The reason that the percentages add up to over 100% is because this part of the survey involved selecting all that applied. One specific pattern that I notice is that sleeping separately “a few times per month” is clearly the least common response among all reasons. Most of the reasons have the highest percentage of sleeping separately “every night”, except for arguments and fights which has by far the lowest percentage. This makes sense to think about because couples do not usually have arguments or fights every night. They generally happen on occasion, which is why the arguments/fights graph has the highest percentage of sleeping separately “once a year or less” and “once a month or less”. Besides arguments and fights, the graphs have similar distributions.

In terms of improvements, majority of couples who sleep in separate beds every night who did not answer neutrally, reported strongly agreeing with the statement “it helped save our relationship”. Most notably, the vast majority of couples sleeping separately every night stated that sleeping separately improved their sleep. When it comes to improvements in sex life, the results are less descriptive and there seems to be less unanimous improvements in this category. Overall, the reasons that couples chose to sleep apart every night were practical sleep reasons relating to sleeping habits such as temperature differences, middle-of-the-night disturbances, snoring, and cover-sharing.

# 3 Machine Learning Methods:

Apply one or more of the modeling or other machine learning techniques that we’ve learned in class. Describe your reasons for using the method – why it is appropriate. Describe the conclusions of each particular analysis. Include more visualizations, as needed.

Can we predict if a couple sleeps separately at all based on Relationship Status, Relationship Length, Age, and Current Occupation?

The feature we want to classify is How\_Often\_Sep\_Beds (how often couples sleep separately). We will be using a decision tree to model the data because each of the features which we are analyzing are categorical variables and the target variable is also categorical. Something to note is that the vast majority of people did answer that they never sleep separately and so we will be adjusting the frequency of sleeping separately responses to be a binary response (No = never sleep separately, Yes = do sleep separately to some degree). This will simplify the response more and allow for a model to be more useful.

## 3.1 Create Training and Test Data

RNGversion('4.0.0')  
set.seed(123)  
  
# Remove empty responses  
sleeping\_sep\_ML <- sleeping\_alone\_data %>%   
 select(How\_Often\_Sep\_Beds, Relationship\_Length, Age) %>%   
   
 filter(!How\_Often\_Sep\_Beds=="", !Age=="", !Relationship\_Length=="") %>%   
   
 # Make the response binary  
 mutate(Binary\_Sleeping\_Sep = factor(How\_Often\_Sep\_Beds,   
 levels = c("Never", "A few times per month", "A few times per week", "Once a year or less", "Once a month or less", "Every night"),  
 labels = c("No", "Yes", "Yes", "Yes", "Yes", "Yes"))) %>%  
   
 dplyr::select(-How\_Often\_Sep\_Beds)   
   
  
holdout\_split <- function(df, pred, train\_percent = 0.80){  
  
 df\_y <- df[,pred]  
 df\_split <- sample.split(df\_y, SplitRatio = train\_percent)  
   
 return(list(train\_x = tibble(df[df\_split, colnames(df)!=pred]),   
 train\_y = df\_y[df\_split],   
 test\_x = tibble(df[!df\_split, colnames(df)!=pred]),   
 test\_y = df\_y[!df\_split]))  
}  
  
# Create training and testing data sets  
holdout\_sleeping <- holdout\_split(df = sleeping\_sep\_ML, pred = "Binary\_Sleeping\_Sep", train\_percent = 0.70)  
  
# Check the proportions  
table(holdout\_sleeping$train\_y) %>%   
 prop.table() %>%   
 round(digits = 5)

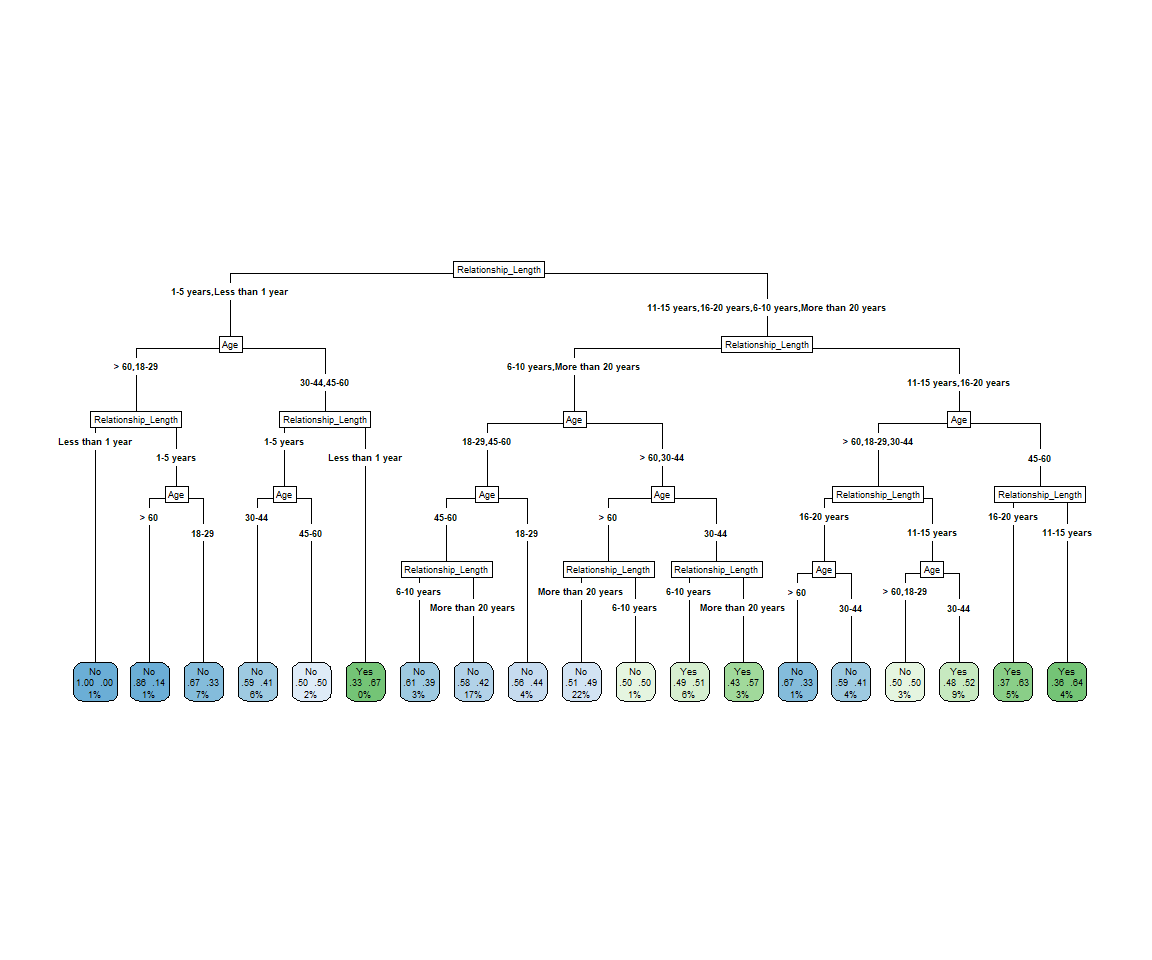
##   
## No Yes   
## 0.53627 0.46373

table(holdout\_sleeping$test\_y) %>%   
 prop.table() %>%   
 round(digits = 5)

##   
## No Yes   
## 0.53642 0.46358

## 3.2 Training a Model on the Data

# Create model and train it on the training data  
train\_df <- data.frame(holdout\_sleeping$train\_x,  
 Binary\_Sleeping\_Sep = holdout\_sleeping$train\_y)  
   
sleeping\_tree <- rpart(formula = Binary\_Sleeping\_Sep ~ .,  
 data = train\_df,  
 method = "class",  
 parms = list(split = "information"),  
 minsplit = 0,   
 minbucket = 0,  
 cp = -1)   
  
rpart.plot(x = sleeping\_tree,  
 type = 5,  
 extra = 104)



## 3.3 Pruning the Tree

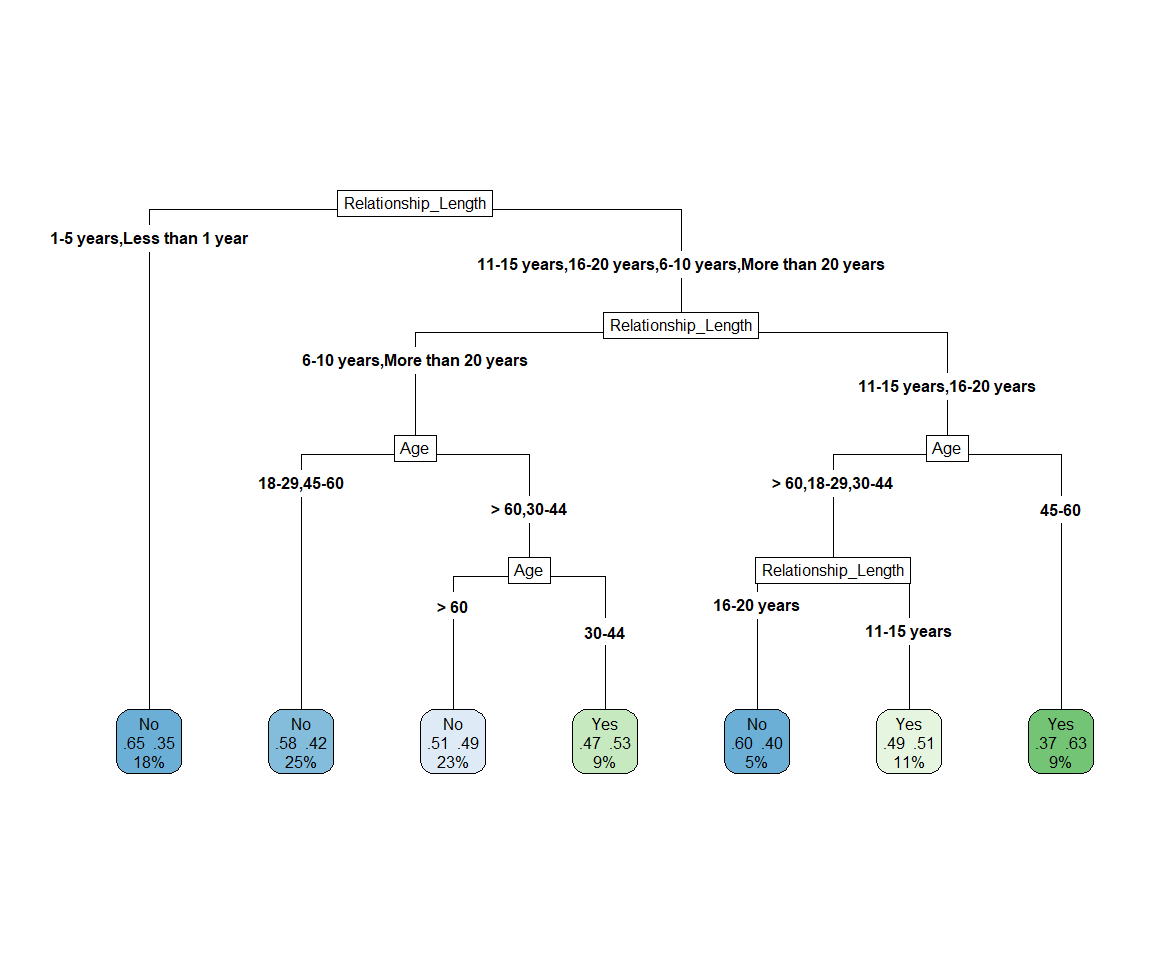
# Finding the lowest xerror:  
sleeping\_tree$cptable %>%   
 data.frame() %>%   
 filter(xerror == min(xerror))

## CP nsplit rel.error xerror xstd  
## 1 0.01840491 0 1 1 0.04055872

# Locating where to prune the tree - simplest tree with xerror < min(xerror) + xstd:  
sleeping\_tree$cptable %>%   
 data.frame() %>%   
 filter(xerror < min(xerror) + 0.04055872)

## CP nsplit rel.error xerror xstd  
## 1 0.01840491 0 1 1 0.04055872

# CP returned is 0.01, go a little below   
sleeping\_pruned <- prune(sleeping\_tree,  
 cp = 0.006)  
  
rpart.plot(x = sleeping\_pruned,  
 type = 5,  
 extra = 104)



## 3.4 Evaluating Model Performance: Confusion Matrix

sleeping\_pred <- predict(object = sleeping\_pruned,  
 newdata = holdout\_sleeping$test\_x,  
 type = "class")  
  
  
# Confusion Matrix of predicted vs actual classes  
cm\_sleeping <-   
 confusionMatrix(data = sleeping\_pred,   
 reference = holdout\_sleeping$test\_y,  
 positive = "No",  
 dnn = c('predicted', 'actual'))  
  
cm\_sleeping

## Confusion Matrix and Statistics  
##   
## actual  
## predicted No Yes  
## No 128 87  
## Yes 34 53  
##   
## Accuracy : 0.5993   
## 95% CI : (0.5417, 0.655)  
## No Information Rate : 0.5364   
## P-Value [Acc > NIR] : 0.0161   
##   
## Kappa : 0.1731   
##   
## Mcnemar's Test P-Value : 2.276e-06   
##   
## Sensitivity : 0.7901   
## Specificity : 0.3786   
## Pos Pred Value : 0.5953   
## Neg Pred Value : 0.6092   
## Prevalence : 0.5364   
## Detection Rate : 0.4238   
## Detection Prevalence : 0.7119   
## Balanced Accuracy : 0.5843   
##   
## 'Positive' Class : No   
##

## 3.5 Model Reflection:

The accuracy of this pruned decision tree model on this data is about 60% which is somewhat low. Considering the two features that were analyzed in relation to frequency sleeping separately and that there were only two features that are fairly general (Relationship length and Age), the model being able to predict if a couple never sleeps separately with a 60% accuracy is not bad. The single CP value that was given was close to 0.01, but after trying 0.008, 0.009, and 0.01, all of these CP values pruned the tree a little too much and so we went with a CP value of 0.006 which resulted in the pruned tree above. One thing that the decision tree illustrated that was an unexpected insight was that people in age groups over 60, 18-29, and 30-44 are shown to be more likely to not sleep separately if they have been in a relationship for 16-20 than if they have been in a relationship 11-15 years.

# 4 Conclusions:

This data does illustrate a tendancy toward stereotypical patterns in our society regarding age / relationship length and sleeping situation. That being that couples who are older (or have one partner who is older) are more likely to sleep in separate beds. Additionally couples who have been together for at least a decade are much more likely to sleep in separate beds. That being said, the purpose of this data was to help shed light on the factors involved in sleeping separately and allow us to examine the reasons why couples choose to sleep separately. It seems that the reasons are highly related to comfort and to getting the best quality of sleep. The stereotype of age / relationship length and separate sleeping usually has to do with stereotypes about “romance fading away” or things of the kind. However, the most popular reasons in general are far more specific and practical. It is understandable that couples who have been together longer (say 16-20) years may be at the point where they can comfortably communicate that it would be best to have separate beds. This data does show some nuance, however, in the relation between age of a person, relationship length, and the likelihood that they sleep separately at all because in ages 44-60 the person is shown to be more likely to sleep separately from their partner than ages over 60. There are more factors that could be considered though. For example, the lens of the data could zoom in more and look specifically at people above 40 where both people in the relationship are above the age of 30 and aim to examine more reasons for sleeping separately that could have more emotional ties than practical ties.

# 5 Limitations / Recommendations:

There are several limitations to this data. The aspect of gender identity is neglected in this survey, only male and female options are given and there is no indication of which gender identity the partner or the person taking the survey is. It would be an interesting aspect to consider the patterns among different types of couples in the US (in terms of gender and sexual identity and how that varies depending on age/generation).