Bivariate Graphing Assignment

Henry Lin

```
library(ggplot2)
library(knitr)
library(dplyr)
library(ggplot2)
library(janitor)
library(purrr)
library(kableExtra)
library(ggmosaic)
library(scales)
opts_chunk$set(echo = TRUE, warning=FALSE, message=FALSE)
```

```
load("/cloud/project/data/clean.Rdata")
load("/cloud/project/data/current_status.Rdata")
load("/cloud/project/data/family_impact_on_work.Rdata")
load("/cloud/project/data/hours_worked.Rdata")
load("/cloud/project/data/job_satisfaction.Rdata")
load("/cloud/project/data/mydata.Rdata")
load("/cloud/project/data/number_of_jobs.Rdata")
load("/cloud/project/data/seperation_cause.Rdata")
load("/cloud/project/data/time_of_day_worked.Rdata")
load("/cloud/project/data/work_impact_family_time.Rdata")
load("/cloud/project/data/addhealth4.RData")
```

Variables of Interest

Original Variable (from addhealth)	New Variable	Categorical or Numerical	Description (from addhealth)
H4LM20	time_of_day_worked	Categorical	Which one of these categories best describes the hours you (work/worked) at this job?
H4LM17	separation_cause	Categorical	What is the main reason you left your most recent job?
H4LM26	job_satisfaction	Categorical	How satisfied (are/were) you with this job?
H4LM14	current_status	Categorical	Which one of the following categories best describes what you're doing now?

Original Variable (from addhealth)	New Variable	Categorical or Numerical	Description (from addhealth)
H4LM28	family_impact_on_work	Categorical	Indicate how much you would agree or disagree with this statement: Family responsibilities have interfered with my ability to work.
H4LM30	work_impact_family_time	Categorical	(In the past 12 months/Since you started your current job/In the last year of your most recent job), how often on your primary job (have you spent/have you spent/did you spend) less time with your family than you wanted to because of work responsibilities?
H4LM19	hours_worked	Numerical	How many hours a week (do/did) you usually work at this job?
H4LM12	number_of_jobs_worked	Numerical	On how many jobs are you currently working for pay at least 10 hours a week?

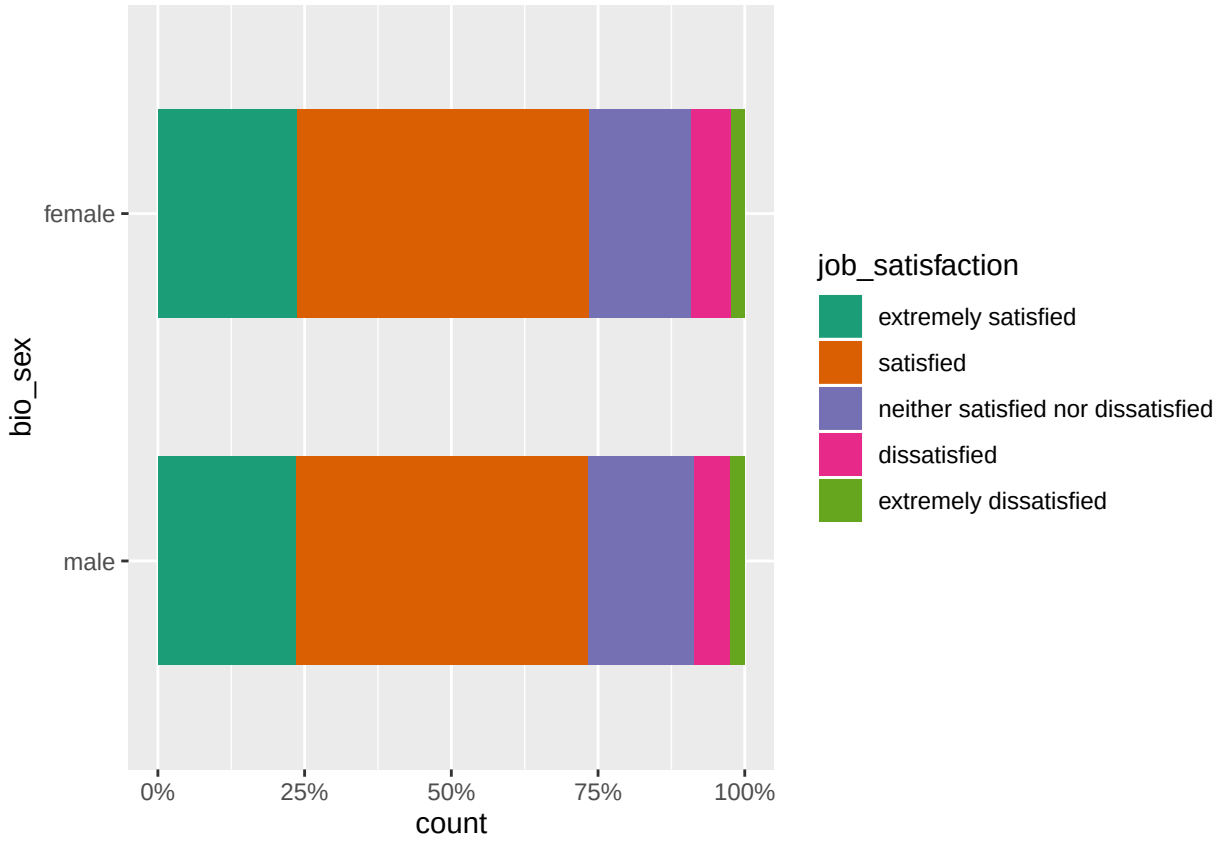
C ~ C Association

time_of_day_worked is the shifts worked by the respondents. job_satisfaction is respondent's satisfaction with their employment.

```
mydata%>%

  filter(job_satisfaction %in% c ("extremely satisfied", "satisfied", "neither satisfied nor dissatisfi

  droplevels() %>%

  tabyl( bio_sex, job_satisfaction) %>%
    adorn_totals(c("row")) %>%
    adorn_percentages("row") %>%
    adorn_pct_formatting(digits=2) %>%

  kable() %>%
    kable_styling(bootstrap_options = "striped",
                  full_width = T,
                  position="left") %>%
  add_header_above(c("","job satisfaction"=5))
```

bio_sex	job satisfaction				
	extremely satisfied	satisfied	neither satisfied nor dissatisfied	dissatisfied	extremely dissatisfied
male	23.47%	49.91%	18.02%	6.14%	2.46%
female	23.80%	49.82%	17.23%	6.90%	2.25%
Total	23.65%	49.86%	17.60%	6.55%	2.35%

```
mydata%>%

  filter(job_satisfaction %in% c ("extremely satisfied", "satisfied", "neither satisfied nor dissatisfi
  droplevels() %>%

  ggplot(aes(x=bio_sex, fill=job_satisfaction))+
  geom_bar(position = position_fill(reverse = T), width=.6) +

  scale_fill_brewer(palette="Dark2") +
  scale_y_continuous(labels=percent) +
  coord_flip()
```



Narrative

When looking at the all addhealth4observations, the rates of job satisfaction are similar for both sexes, showing no potential association. When comparing job satisfaction with gender, 23.47% of men responsed with extremely satisfied, compared to 23.80% for women, for difference in sample proportions of 0.33%. 49.91% of men responsed with satisfied, compared to 49.82% for women, for difference in sample proportions of 0.09%. 18.02% of men responsed with neither satisfied nor dissatisfied, compared to 17.23% for women, for

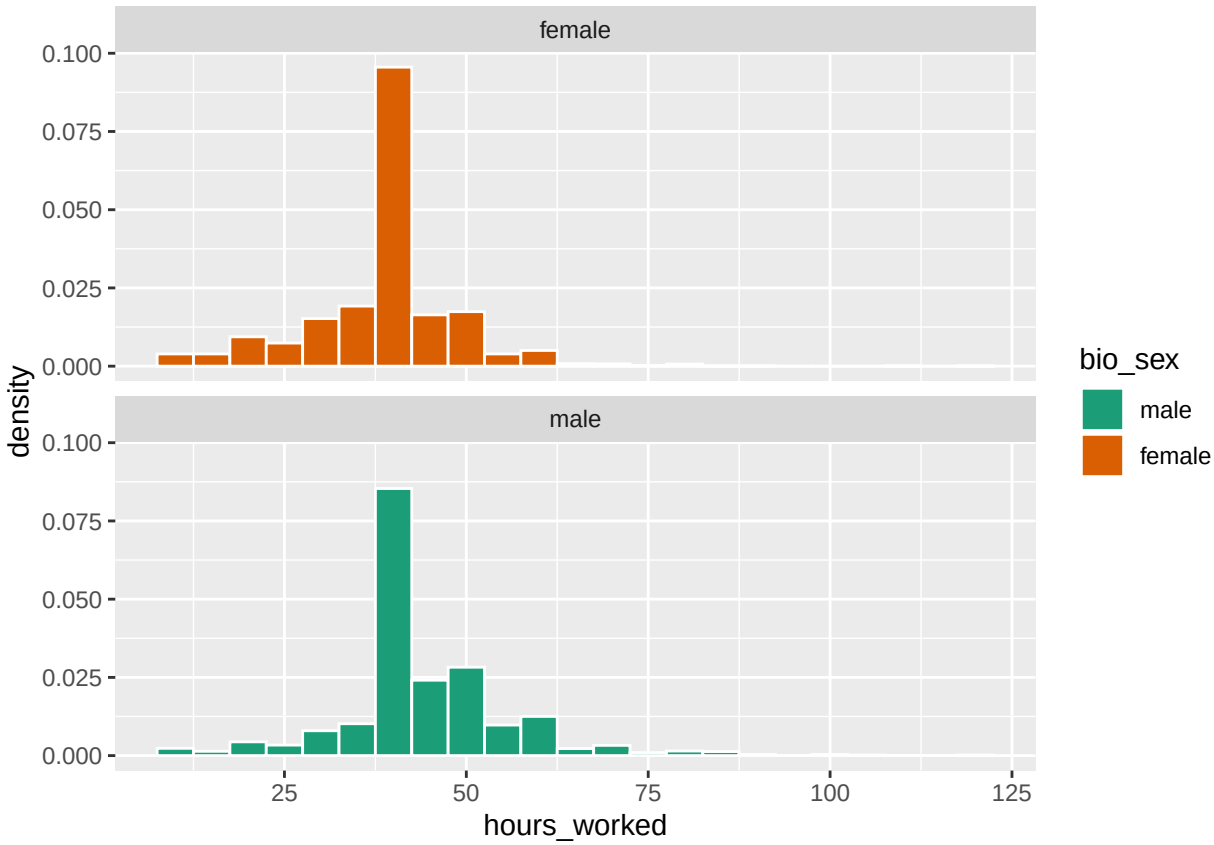
difference in sample proportions of 0.79%. 6.14% of men responsed with dissatisfied, compared to 6.90% for women, for difference in sample proportions of 0.76% 2.46% of men responsed with extremely dissatisfied, compared to 2.25% for women, for difference in sample proportions of 0.21%.
The relationship between sex and job satisfaction was similar for men and women.

Q ~ C Association

```
mydata %>%

  filter(hours_worked < 121) %>%

  group_by(bio_sex) %>%

  summarize(n = n(),
            min = min(hours_worked),
            q1 = quantile(hours_worked, probs=0.25),
            median = median(hours_worked),
            mean = round(mean(hours_worked),1),
            q3 = quantile(hours_worked, probs=0.75),
            max = max(hours_worked),
            sd = round(sd(hours_worked),1)) %>%

  kable() %>%
    kable_styling(bootstrap_options = "striped",
                  full_width = T,
                  position="left")   %>%
    add_header_above(c("hours worked "= 9))
```

	hours worked							
bio_sex	n	min	q1	median	mean	q3	max	sd
male	2315	10	40	40	44.0	50	120	12.0
female	2708	10	35	40	38.8	40	120	10.3

```
mydata %>%

  filter(hours_worked < 121 ) %>%

  ggplot(aes(hours_worked))+

    geom_histogram(aes(y=..density..,
                       fill=bio_sex),
                   color="white",
                   binwidth = 5) +

      scale_fill_brewer(palette="Dark2") +

    facet_wrap(~ bio_sex,
               ncol=1,
               as.table = FALSE)  ## reverses the order of the levels to match boxplots
```



Calculate grouped summary statistics (means, sd by group) and compare them using sentences with numbers to justify your claims. Compare center, shape and spread of the quantitative variable across different levels of the categorical variable.

Although both males and females work a median of 40 hours and worked the same minimum and maximum hours, 10 hours and 120 hours respectively, males work a mean of 44 hours, compared to women, at 38.8 hours. This represents a difference of 13.40% of working hours for males compared to females. There are further differences in addition. The interquartile range for males is higher at 40 to 50 hours, compared to women at 35 to 40 hours.

(B ~ C, C ~ B, or Q ~ B)

same format as above

Q ~ Q Association

Calculate the correlation and use it when discussing the direction, strength, and form of this relationship.