Hmk4-Q3

Hrishikesh Telang (hnt2107)

11/14/2018

The data were obtained in a survey of students math and portuguese language courses in secondary school. It contains a lot of interesting social, gender and study information about students.

Attributes used:

sex - student's sex (binary: 'F' - female or 'M' - male)

age - student's age (numeric: from 15 to 22)

activities - extra-curricular activities (binary: yes or no)

higher - wants to take higher education (binary: yes or no)

goout - frequency of going out with friends (numeric: from 1 - very low to 5 - very high)

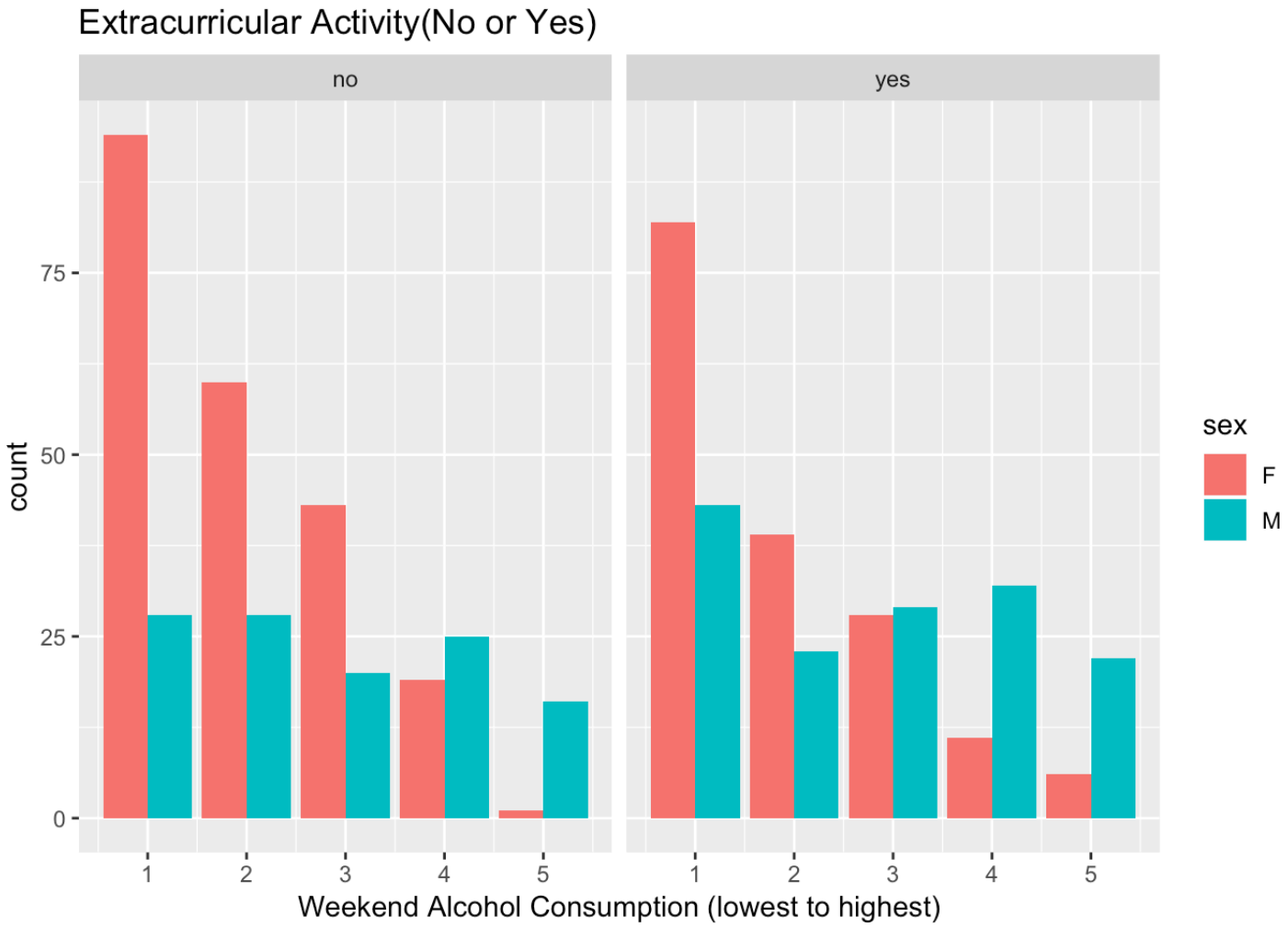
Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)

```
library(tidyverse)
library(dplyr)
library(plotly)
library(ggplot2)
library(readr)

alc<-read_csv('/Users/hrishikeshtelang/Desktop/DSI acads/EDAV/student-por.csv')
alc
```

```
## # A tibble: 649 x 33
##    school sex     age address famsize Pstatus  Medu  Fedu Mjob  Fjob
##    <chr>  <chr> <int> <chr>   <chr>   <chr>   <int> <int> <chr> <chr>
##  1 GP     F        18 U       GT3     A           4     4 at_h… teac…
##  2 GP     F        17 U       GT3     T           1     1 at_h… other
##  3 GP     F        15 U       LE3     T           1     1 at_h… other
##  4 GP     F        15 U       GT3     T           4     2 heal… serv…
##  5 GP     F        16 U       GT3     T           3     3 other other
##  6 GP     M        16 U       LE3     T           4     3 serv… other
##  7 GP     M        16 U       LE3     T           2     2 other other
##  8 GP     F        17 U       GT3     A           4     4 other teac…
##  9 GP     M        15 U       LE3     A           3     2 serv… other
## 10 GP     M        15 U       GT3     T           3     4 other other
## # ... with 639 more rows, and 23 more variables: reason <chr>,
## #   guardian <chr>, traveltime <int>, studytime <int>, failures <int>,
## #   schoolsup <chr>, famsup <chr>, paid <chr>, activities <chr>,
## #   nursery <chr>, higher <chr>, internet <chr>, romantic <chr>,
## #   famrel <int>, freetime <int>, goout <int>, Dalc <int>, Walc <int>,
## #   health <int>, absences <int>, G1 <int>, G2 <int>, G3 <int>
```

```
g <- ggplot(alc, aes(x =Walc, fill=sex))+geom_bar(position = "dodge")+facet_wrap(~act
ivities)+xlab("Weekend Alcohol Consumption (lowest to highest)")+ggtitle("Extracurric
ular Activity(No or Yes)")
g
```



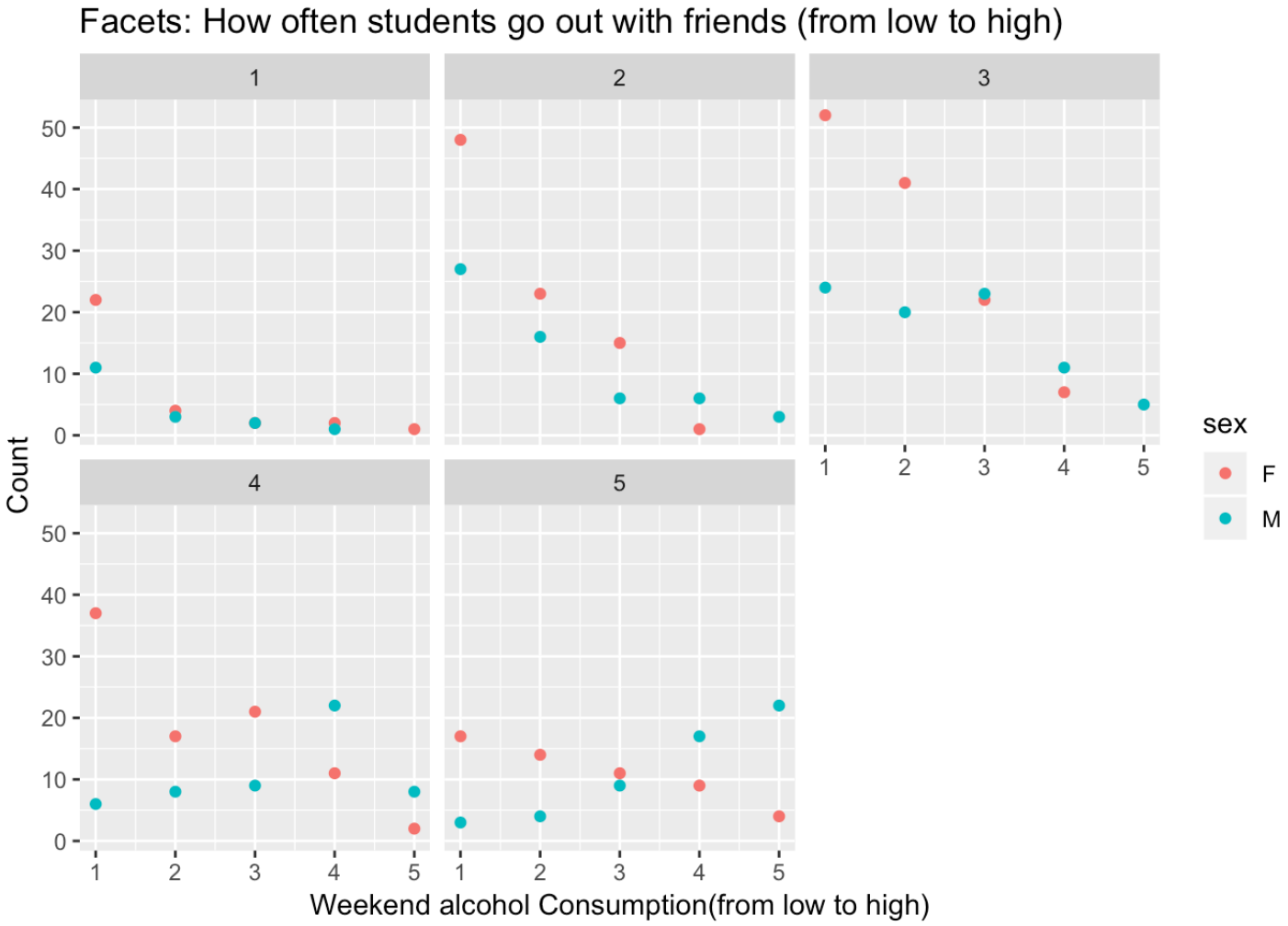
The variable chosen is ‘Walc’ or Weekend Alcohol Consumption.

The first vizualization is a bar graph that vizualizes the total number of students by each level of alcohol consumption(1-lowest 5-highest) distributed on the basis of gender, and faceted on whether or not they engage in extracurricular activities.

1. A consistent pattern observed is that the number of females indulging in heavy drinking goes down regardless of extracurricular activities.
2. However, an anomaly is that there is a greater number of females with extracurricular activities having a value of 5 (high) drinking compared to females who don’t have activities.
3. The number of males doesn’t appear to follow a decreasing pattern that is observed in females. However, more men having extracurricular activities indluge in heavy drinking on weekends compared to men who don’t have extracurricular activities. This is an interesting observation and a potentially important question, as students with activities usually don’t get a lot of free time.

```
counts2 <-alc %>% group_by(Walc,goout,sex) %>% summarize(Freq = n())

ggplot(counts2, aes(x = Walc, y = Freq, color =sex)) +
  geom_point()+facet_wrap(~goout)+xlab("Weekend alcohol Consumption(from low to high)
")+ylab("Count")+ggtitle("Facets: How often students go out with friends (from low to
high)")
```



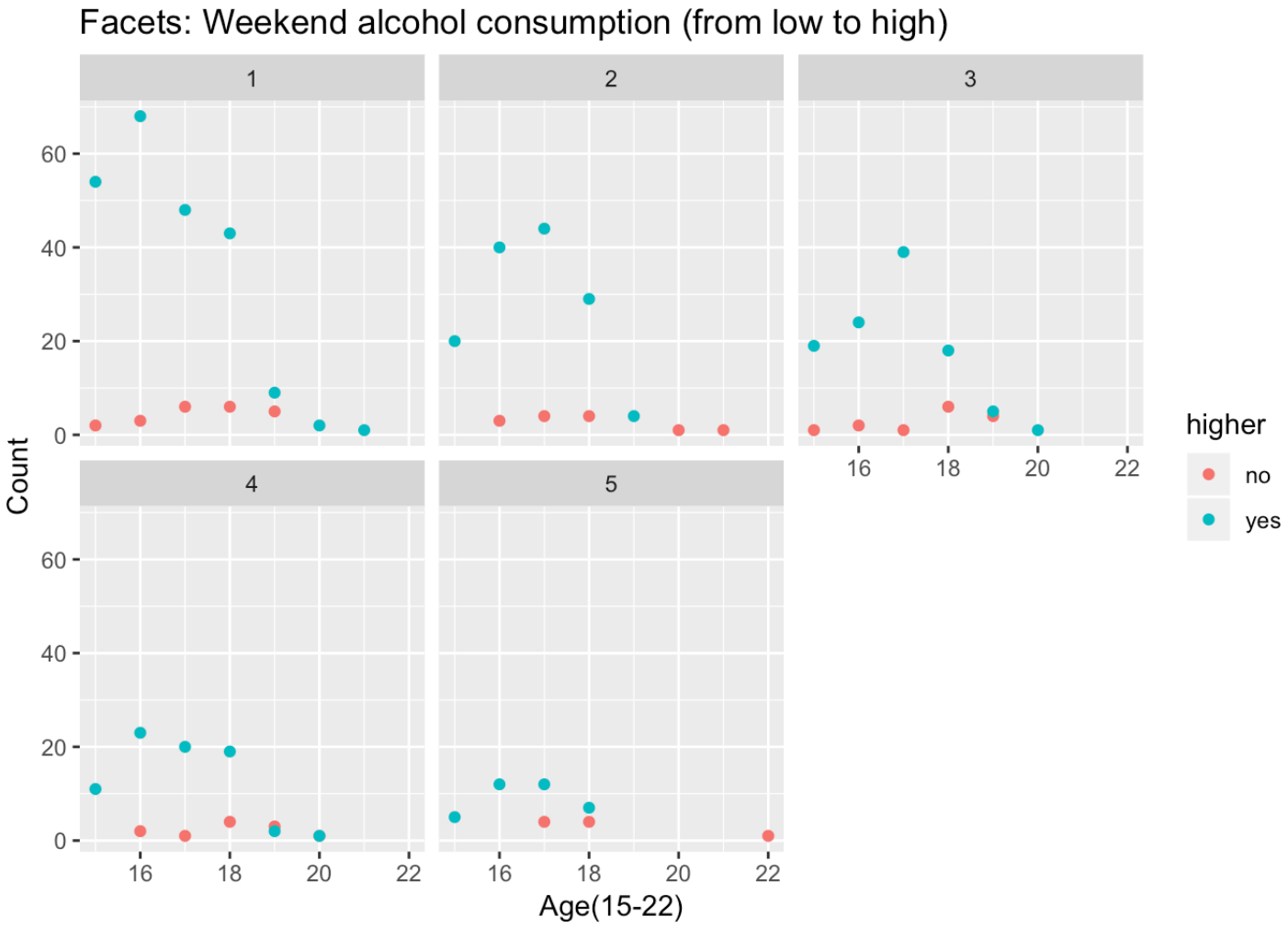
The second visualization based on 'Walc' plots its distribution with how often students go out with friends again divided based on gender.

1. The general trend among people who don't go out often is that they are light to moderate drinkers. In fact, females almost never indulge in heavy (level 4-5) drinking regardless of whether or not they go out often.

2. An interesting insight gained is that males who go out very frequently (level 5), almost never engage in light drinking, and in fact have higher number of heavy drinkers, bucking the general trend.

```
counts1 <-alc %>% group_by(Walc,age,higher) %>% summarize(Freq = n())

ggplot(counts1, aes(age, Freq, color = higher)) + geom_point()+facet_wrap(~Walc)+xlab
("Age(15-22)")+ylab("Count")+ggtitle("Facets: Weekend alcohol consumption (from low t
o high)")
```



The third visualization denotes number of students interested in pursuing higher education by age, faceted by level of weekend drinking

1. It is observed that students who do want to pursue higher education have the most number of light drinkers (levels 1-2). There are very few heavy drinkers interested in pursuing higher education between the ages 15-22