Brian Lambert
STA 404
Module 4 Homework

1. **Most Female and Male Occupations in 2015**

Larger circles represent more common jobs.

MORE MALE   EVEN   MORE FEMALE

CONSTRUCTION   EDUCATION AND LIBRARY
Carpenters   Preschool and
                                                     Kindergarten Teachers

0%   10%   20%   30%   40%   50%   60%   70%   80%   90%   100%

PERCENTAGE WHO ARE FEMALE
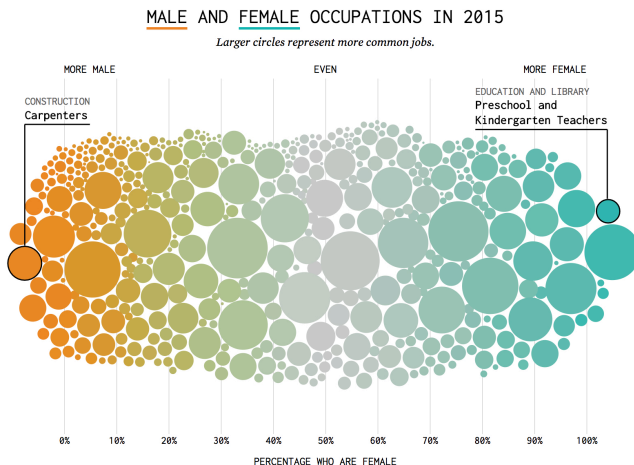
   a. **Evaluate this graphic in terms of what preattentive attribute(s)**

   This graphic is oriented entirely on the x-axis, which is the percentage of females for a specific job. The y-axis has no significance in this display besides providing space for a clean display of all data points. The shape is a cloud of data plotted on top of a rectangular, 2-D plot. The vertical lines showing the x-axis values extend the vertically through the entire plot, however, the lines are thin and a light grey color to avoid drawing the audience's attention away from the real data (the circles for each job). The graphic also includes various added marks. First, the creator included endpoints on the top sides of the graphic to further explain to the user that the x-axis displays percentage of females in the jobs, increasing from left to right. Another added mark was identifying on the leading jobs with the least women and another mark for one with the most women. Finally, the author includes a subtitle that explains that the size of each circle corresponds to how common the job is. The graphic also relies on hue to show difference in percentage of females. Male intensive jobs have an orange color, even sex jobs a grey color, and female intensive jobs have a teal hue. Color intensity is then used to show higher and lower percentages of men and women, darker intensities of orange showing less women and darker intensity of teal to show more women. Finally, the graphic uses a dodging or jittered spatial position so that each data point is visible / hoverable to show the title of each job.

   b. **Comment on the use of color in this display**

The use of a color gradient was not necessary in this display as the a-axis already shows the percentage of women in each job, however, I think the inclusion of coloring was a great idea because it further drives home the point of the division of labor. The creator likely includes color in this way because he/she wants us to examine the specific jobs that are heavy with men, heavy with women, and even between both sexes. The graphic also cleverly uses this color scheme by underlining "men" in the title with the orange color while "women" is underlined with the teal color. This was a clever way of subliminally telling the audience what they should be looking at in the graphic.

2. **"Diamonds" data set visualizations**

   Code:

   ```
   # Author: Brian Lambert

   # Name: module_4_homework_code.R

   # Description: Graphics exploring the cut of diamonds in the "diamonds" data set,

   # as well as comparing the cut to diamond size (carats).


   # setwd("/Users/brianlambert/Desktop/STA404/Module_4")

   library(tidyverse)



   #========================= Data =========================


   myDiamonds <- diamonds %>%

      mutate(Gcarat = cut(carat, breaks=seq(from=0,to=3.5,by=.25)))
   ```

```
#========================== Graphic 1 ==========================


# displaying diamond cut by count with a veritcal bar graph and a gradient color
scale

ggplot(myDiamonds) +

    geom_bar(aes(cut, fill = cut)) +

    scale_fill_brewer(palette = "RdYlGn") +

    labs(caption = "Data source: ggplot2 diamonds dataset") +

    theme(panel.grid.major = element_blank(),

        panel.grid.minor = element_blank(),

        panel.border = element_blank(),

        panel.background = element_blank())




#========================== Graphic 2 ==========================


# displaying diamond cut by count with a horizontal bar graph and a gradient color
scale

ggplot(myDiamonds) +

    geom_bar(aes(cut, fill = cut)) +

    scale_fill_brewer(palette = "RdYlGn") +

    theme(panel.grid.major = element_blank(),

        panel.grid.minor = element_blank(),

        panel.border = element_blank(),
```

```
        panel.background = element_blank()) +

    labs(caption = "Data source: ggplot2 diamonds dataset") +

    coord_flip()
```

#========================= Graphic 3 =========================

```
# pie chart displaying the counts of eacch diamond cut type

# source for position = "fill": https://stackoverflow.com/questions/31165823/ggplot-
making-a-descriptive-bar-graph-with-no-clear-y-variable

ggplot(myDiamonds, aes(x=factor("1"), fill = cut)) +

    geom_bar(position="fill", width=1, ) +

    scale_fill_brewer(type="qual", palette="RdYlGn") +

    coord_polar(theta="y") +

    theme_minimal() +

    theme(axis.title.x = element_blank(),

        axis.title.y = element_blank(),

        legend.title = element_blank(),

        axis.ticks.y = element_blank(),

        axis.text = element_blank()) +

    ggtitle("Diamond Cut Qualities") +

    labs(caption = "Data source: ggplot2 diamonds dataset") +

    theme(plot.title = element_text(hjust = 0.5))
```
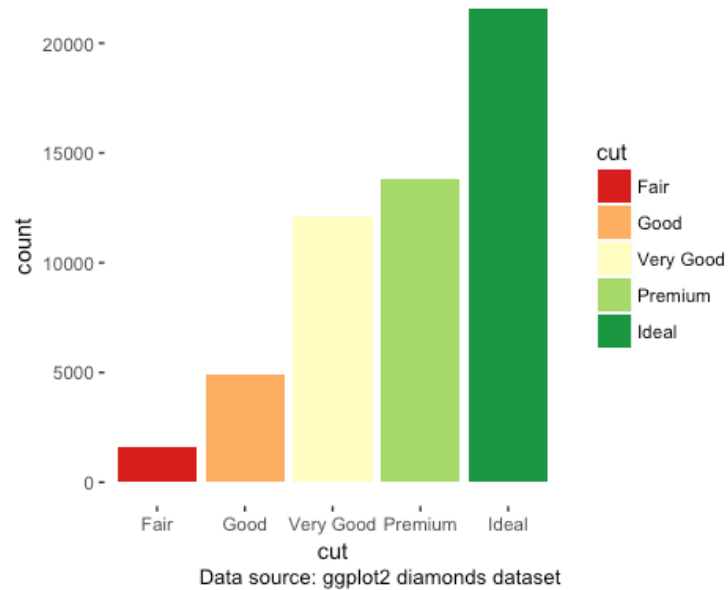
```
#========================= Graphic 4 ==========================


# pie charts  displaying counts of each diamond cut type

# with facet wrap grouped by Gcarat.

ggplot(myDiamonds, aes(x=factor("1"), fill = cut)) +

    geom_bar(position="fill", width=1, ) +

    scale_fill_brewer(type="qual", palette="RdYlGn") +

    coord_polar(theta="y") +

    theme_minimal() +

    theme(axis.title.x = element_blank(),

        axis.title.y = element_blank(),

        legend.title = element_blank(),

        axis.ticks.y = element_blank(),

        axis.text = element_blank()) +

    ggtitle("Diamond Cut Qualities") +

    theme(plot.title = element_text(hjust = 0.5)) +

    labs(caption = "Data source: ggplot2 diamonds dataset") +

    facet_wrap(~Gcarat)
```
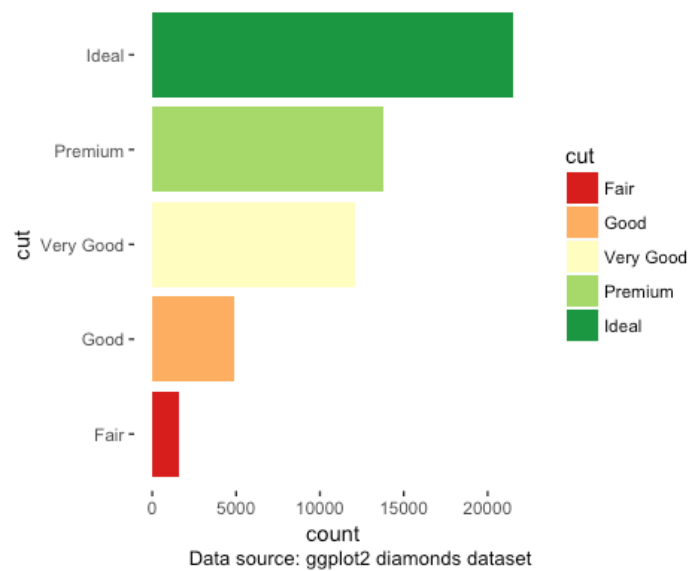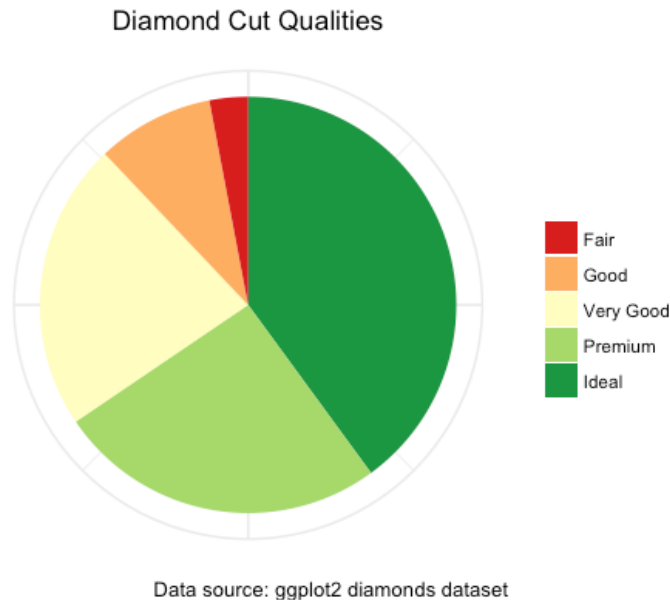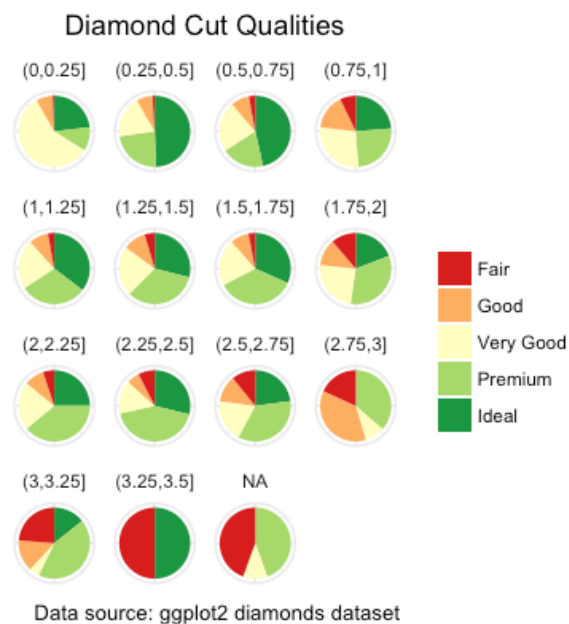
a. vertical bar chart of the cut of diamonds

Data source: ggplot2 diamonds dataset

b. horizontal bar graph of the cut

Data source: ggplot2 diamonds dataset

c. pie chart of the cut

Diamond Cut Qualities

Data source: ggplot2 diamonds dataset

d. compare how the mix of cut varies with different size diamonds

Diamond Cut Qualities

Data source: ggplot2 diamonds dataset

e. conclusion about how cut quality changes with diamond size (carat)

It is easily to see in our final graphic that diamond cut quality has a negative correlation with diamond size. This can be seen as the proportion of diamonds with a fair (the worst) cut steadily increases as diamond size increase. The part of this graphic that might confuse the audience on the true trend is that the proportion of ideal cuts stays relatively constant in each plot. However, one must look at the mean of each diamond cut for each grouping. When viewing the

data with this is mind, it is easy to see that the mean diamond cut quality decreases as the diamond size increases.