Problem Set 2

ylelkes

September 11, 2015

1. Fix each of the following common data frame subsetting errors:

```
mtcars[mtcars$cyl = 4, ]
mtcars[-1:4, ]
mtcars[mtcars$cyl <= 5]
mtcars[mtcars$cyl == 4 | 6, ]
```

2.

Using the following data frame:

```
data(Ecdat::Caschool)
```

- Calculate the total score of readscr and mathscr and add that total as a new variable to the dataframe. Try to do so in two different ways:
- Using simple addition
- Use one of the functions we learned in class
- Rename the columns in the dataframe
- Subset the data to only include the test scores for Butte county, and keep only the last three columns. Do this twice. Once using indexing and once using subsetting.
- On the original dataset, plot the total test score by the average district income using a scatterplot, change the size of the points to vary by the number of students in the district. Do some research, and change the x-axis, y-axis, and legend title names labels and give them something more sensible. Finally add a blue linear regression line (hint, use geom_smooth()). I personally think the generic background of ggplot is kind of ugly. Change the theme to be a bit more palatable.
- use ggsave to save it as a pdf.

3. Plot the following:

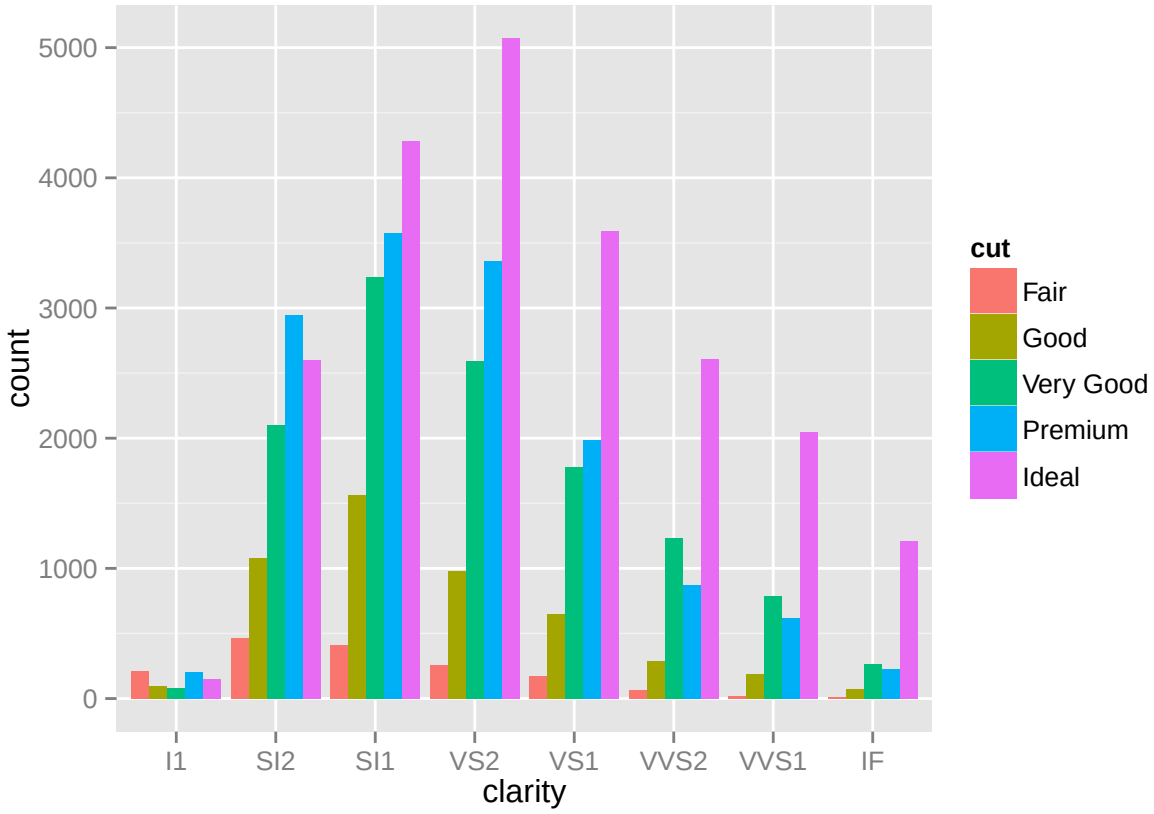
```
plot(c(100, 250), c(300, 450), type = "n", xlab = "", ylab = "")
image <- as.raster(matrix(0:1, ncol = 5, nrow = 3))
rasterImage(image, 100, 300, 150, 350, interpolate = FALSE)
```

- In a single command, change all the black squares to green. (hint, you'll have to look up the hex color code for green)

4.

Using the dataset diamonds (which doesn't need to be loaded), recreate this barplot (not histogram)

```
## Warning: package 'ggplot2' was built under R version 3.1.3
```



5.

Create this list of dataframes:

```
list(mtcars,diamonds,mpg)
```

- From the list, calculate the the correlation between miles per gallon in a city and miles per gallon on the highway (from mpg), add that correlation as a new object in the list.
- From within the list, subset the diamonds dataset so it only includes 1 carat or higher diamonds. Replace the diamonds dataset object in the list with the subsetted dataset.

6. One of the most important functions you need to do as a data analyst is recode variables. There are many ways to do so.

- Use the recode command in the car package to recode the number of cylinders in the mtcars so that the numbers are spelled out (e.g., 4 would be "four")
- do the same thing, but this time, do so using indexing, i.e., with square brackets.