# Assignment 7 – R bar graphs and scatter plots

Professor John Sokol | Due 4/19

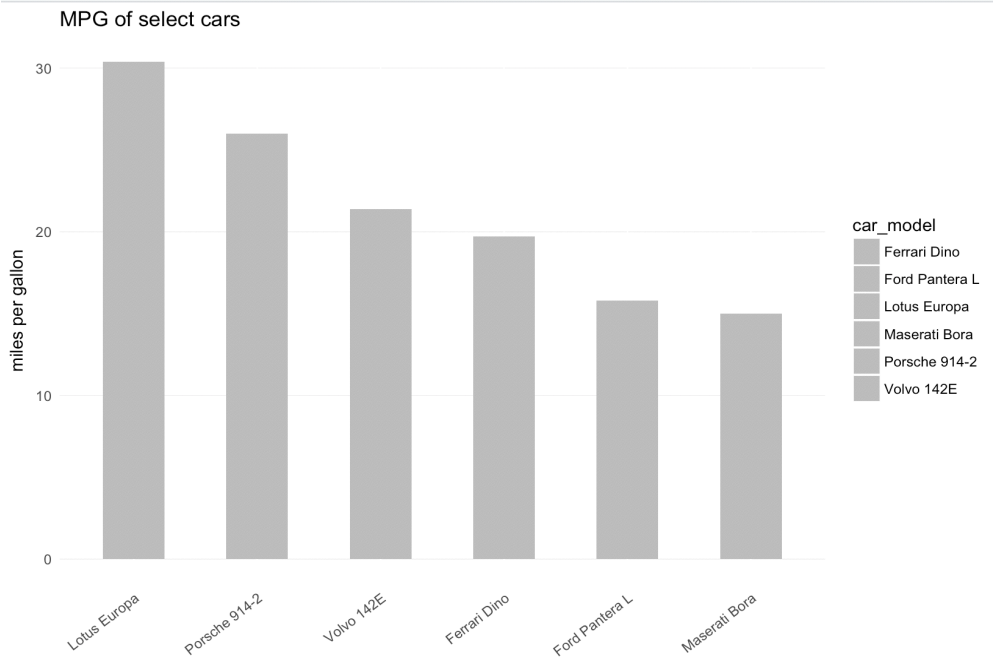
**Rstudio bar graphs:**

Although the code for creating bar graphs looks similar to creating line graphs in R, bar graphs are more nuanced due to the following:

- Different ways to fill bars (count, exact data, summary data)
- Manual color fill in
- Horizontal or vertical bar graph layout
- Graph legend

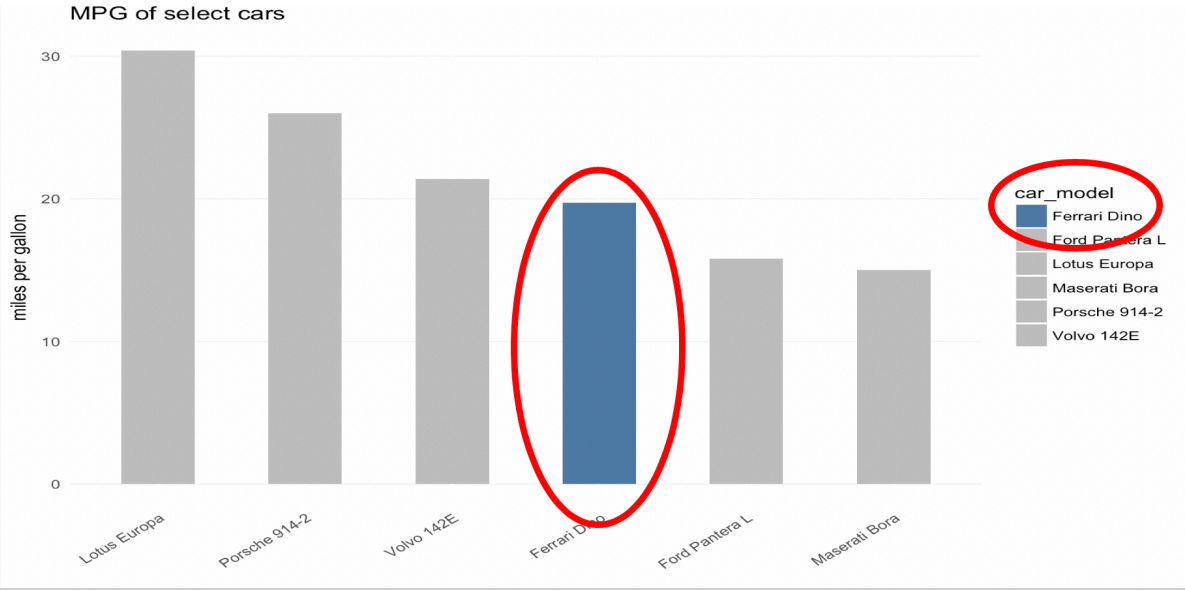
Bar graph of the average MPG's of the last 6 car models in the mtcars DataFrame:

```
library(ggplot2)
ggplot(data = mtcars_bar, aes(x = reorder(car_model, -car_mpg), y = car_mpg, fill = car_model)) + geom_bar(stat = "identity", width = 0.5) + scale_fill_manual(values = c("gray","gray", "gray", "gray", "gray","gray", "gray", "gray")) + theme(panel.background = element_blank(), panel.grid.major.y = element_line(size=.1, color="#DCDCDC"), axis.ticks = element_blank(), axis.text.x = element_text(angle = 38, hjust = 1)) + xlab("") + ylab("miles per gallon") + ggtitle("MPG of select cars")
```



Notice that the bars in the above bar graph are all gray. Observe what happens when a dark blue color is added to the first argument in the scale_manual_color( ) option:

Bar graph of mtcars by MPG




```
scale_fill_manual(values = c("#4E79A7","gray", "gray", "gray", "gray","gray"))
```

Can you identify the pattern to determine how the scale_fill_manual( ) option works?

Fundamental bar graph best practices that were reviewed during the Tableau bar graph lecture are below for reference:

**Bar graphs:**

The bar graph is one of the most important tools in the data visualization toolbox. The human brain can easily compare height, resulting in fast and simple interpretation of what the data represents, so it is important to understand how to create compelling and easily readable bar graph visualizations.

Bar graph formatting guide:

- Ensure data types of data fields are correctly assigned
- Remove excess white space by shrinking y-axis interval
- Employ color to pertinent categories
- Add axis labels as necessary
- Title that states a call to action
- Remove or light gray gridlines
- Ensure data types of data fields are correctly assigned
- Decrease bar height due to low variance (difference in bar height) between each bar
- Change decimals to percent via Default Properties > Number Format
- Add title, remove redundant Location header
- Axis rulers for both rows and columns are set to 'none'
- Columns: Zero lines and grid lines are set to 'none'
- Decrease slightly graph horizontal size
- Increase slightly graph vertical size

Scatter plots:

Scatter plots are simpler compared to bar graphs. Here is the code for a scatter plot in R studio:

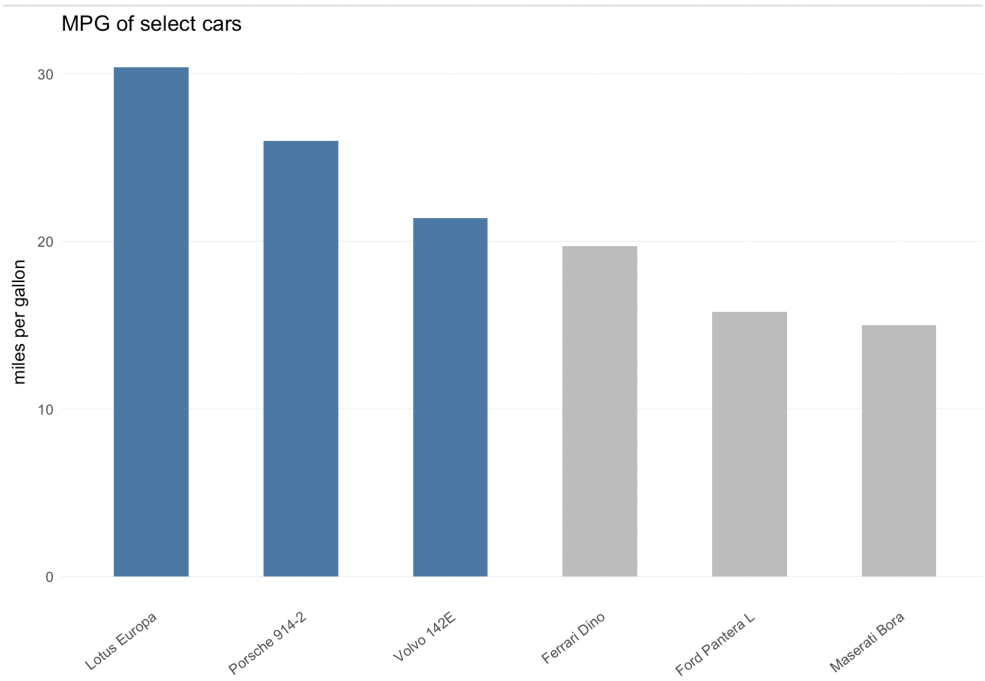
ggplot(data = dataset, aes(x = x_axis , total, y = y_axis)) + geom_point() + theme(panel.background = element_blank(), panel.grid.major.y = element_line(size=.1, color="#DCDCDC"), axis.ticks = element_blank()) + xlab(”x label") + ylab(”y label") + ggtitle(”graph title") + geom_smooth(method=lm)

See the scatter plot visualization output of this code in the deliverables section below.

**Deliverables:**

Bar graphs:

- Submit an R script and a screenshot, PNG or JPEG image of the MPG of select cars bar graph, where the Lotus Europa, Porsche 914-2, and Volvo 142E have a dark blue filled color (#4E79A7) to their bars, and all bars of the other car categories are filled gray. Additionally, remove the default legend with the *legend.position = "none"* option added in theme( ). The bar graph should look like to this:



- Submit a half page to one page write up describing the differences in code between the IT tickets vertical and horizontal bar graphs:

  Vertical bar graph code:

  - ggplot(data = IT_tickets, aes(x = reorder(Location, -ticket_ratio), y = ticket_ratio, fill = Location)) + geom_bar(stat = "summary", fun.y = "mean", width = 0.5) + scale_fill_manual(values = c("gray","red", "gray", "red", "red","gray", "gray", "gray")) + theme(panel.background = element_blank(), panel.grid.major.y = element_line(size=.1, color="#DCDCDC"), axis.ticks = element_blank(), axis.text.x = element_text(angle = 38, hjust = 1), legend.position = "none") + xlab("") + ylab("% ticket ratio") + ggtitle("IT tickets processed by location")

  The difference between the vertical and horizontal code are highlighted in red:

  Horizontal bar graph code:

  - ggplot(data = IT_tickets, aes(x = reorder(Location, -ticket_ratio), y = ticket_ratio, fill = Location)) + geom_bar(stat = "summary", fun.y = "mean", width = 0.5) + scale_fill_manual(values = c("gray","red", "gray", "red", "red","gray", "gray", "gray")) + theme(panel.background = element_blank(), panel.grid.major.x = element_line(size=.1, color="#DCDCDC"), axis.ticks = element_blank(), axis.text.x = element_text(hjust = 1), legend.position = "none") + xlab("") + ylab("% ticket ratio") + ggtitle("IT tickets processed by location") + coord_flip()

Describe what each of the three above red highlighted options do to the graph. Run each code block in Rstudio, then compare and contrast the horizontal and vertical bar graph outputs. Another good troubleshooting strategy is removing just one code option, then run the entire code block to see what visually changed in the visualization to determine what the code actually does.

Scatter plots:

- Create a scatter plot of any two measures in your final project dataset including a trend line. This can also be considered your scatter plot requirement for your final project, so you have the opportunity to knock out two birds with one stone. Find the scatter plot code in the week 12 lecture PowerPoint slides and an image of a scatter plot of four-year graduation rate vs. bachelor degrees awarded within four years below: