

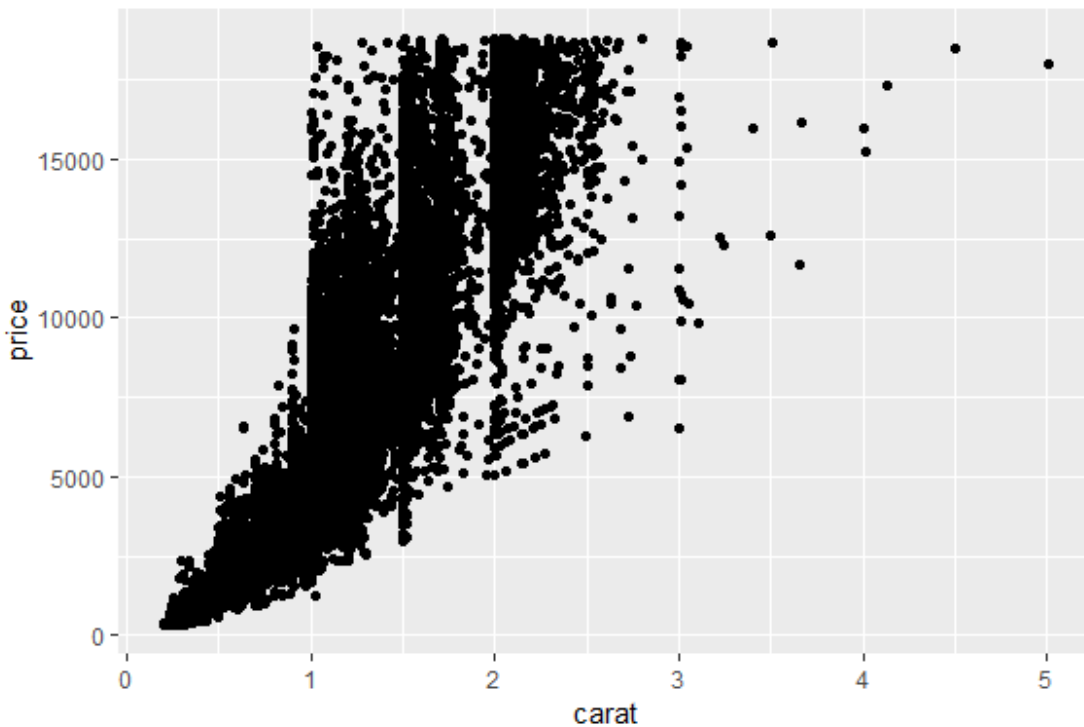
Data Visualization Unit Summary

Yogindra Raghav

9/28/2018

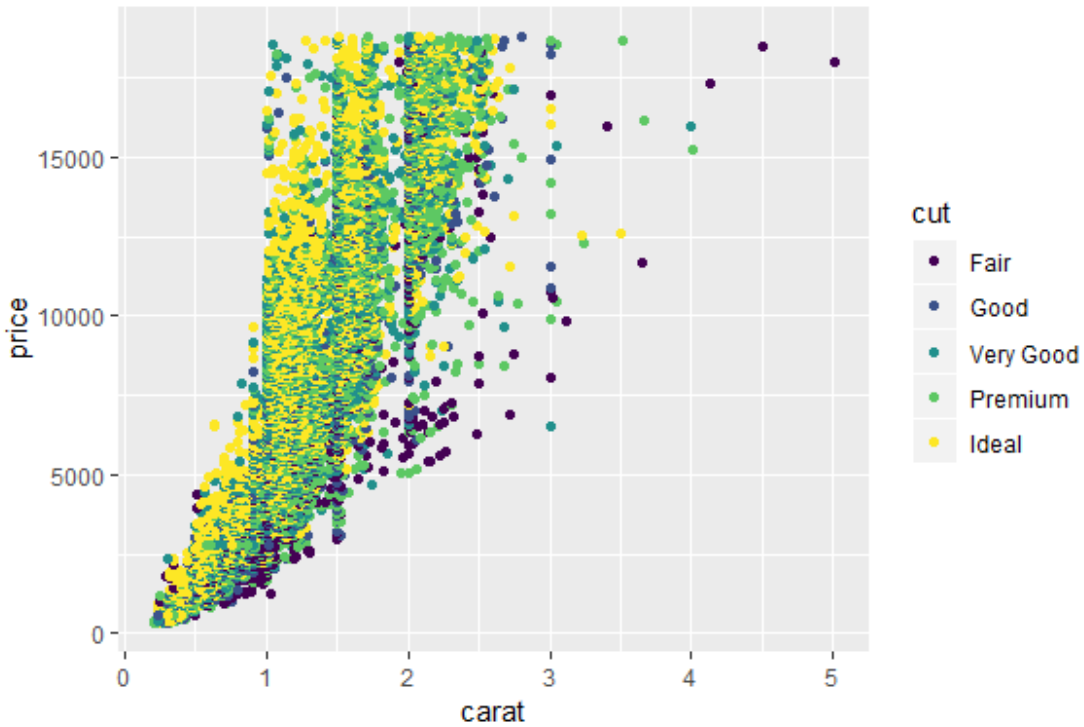
Using the diamonds data set from the ggplot2 package, create a scatter plot of the price (as response) and carat (as explanatory variable)

```
library(ggplot2)
ggplot(data = diamonds) + geom_point(mapping = aes(x = carat, y = price))
```



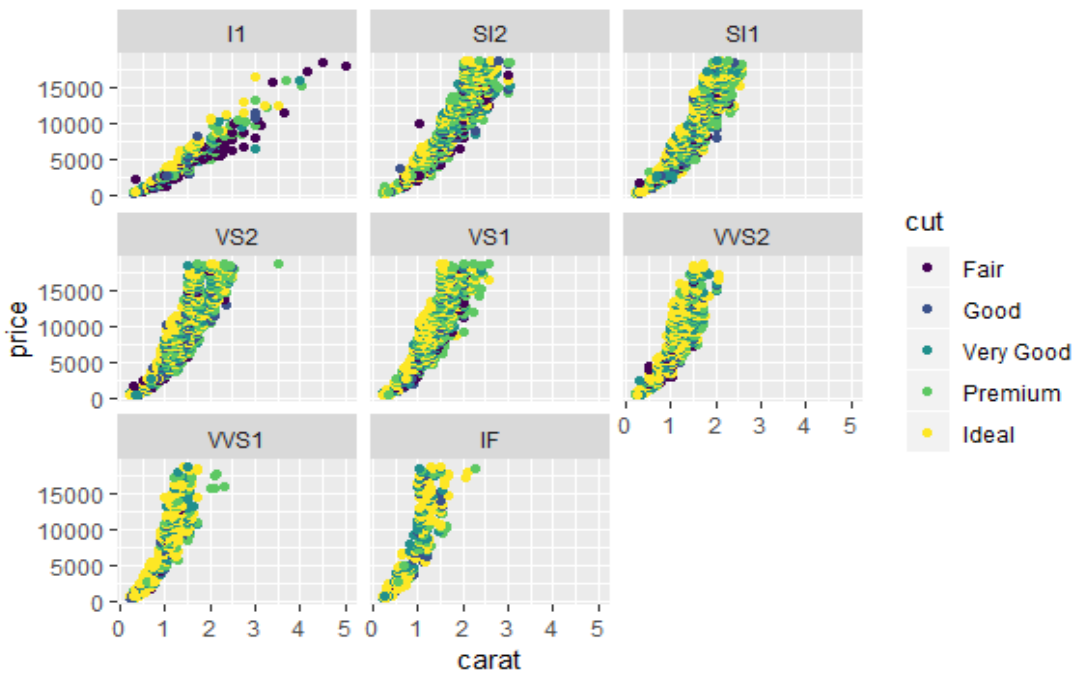
Add another variable, cut, to the scatter plot using color cue.

```
ggplot(data = diamonds) + geom_point(mapping = aes(x = carat, y = price, color = cut))
```



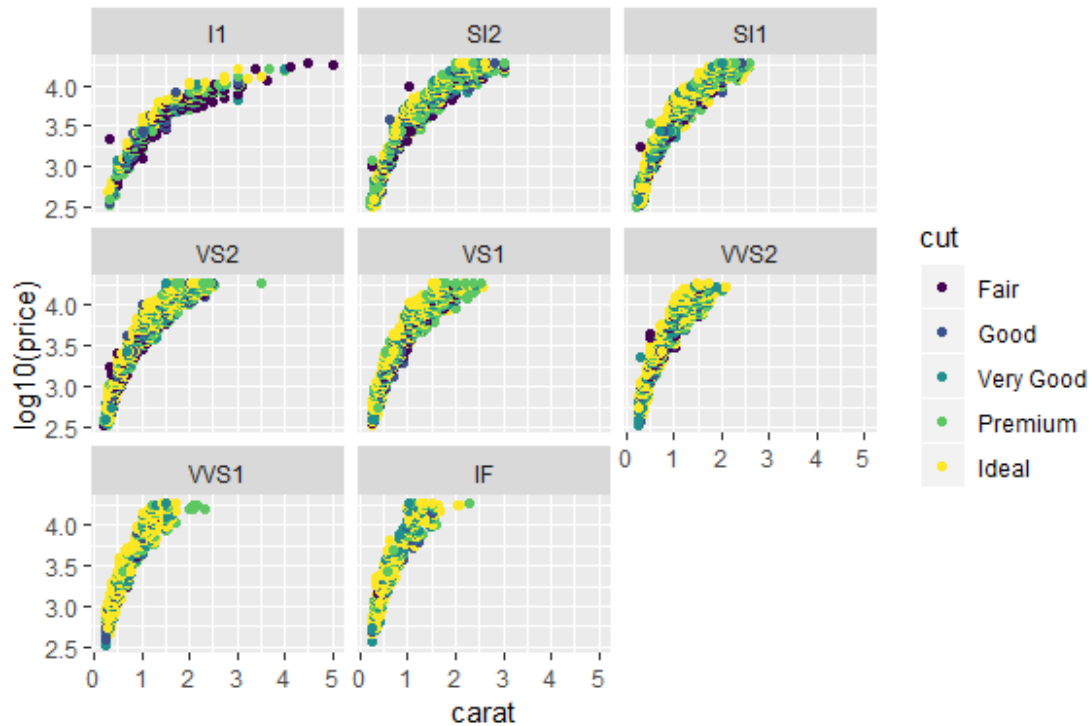
Add yet another variable, clarity, to the scatter plot using facets.

```
ggplot(data = diamonds) + geom_point(mapping = aes(x = carat, y = price, color = cut)) + facet_wrap(~clarity)
```



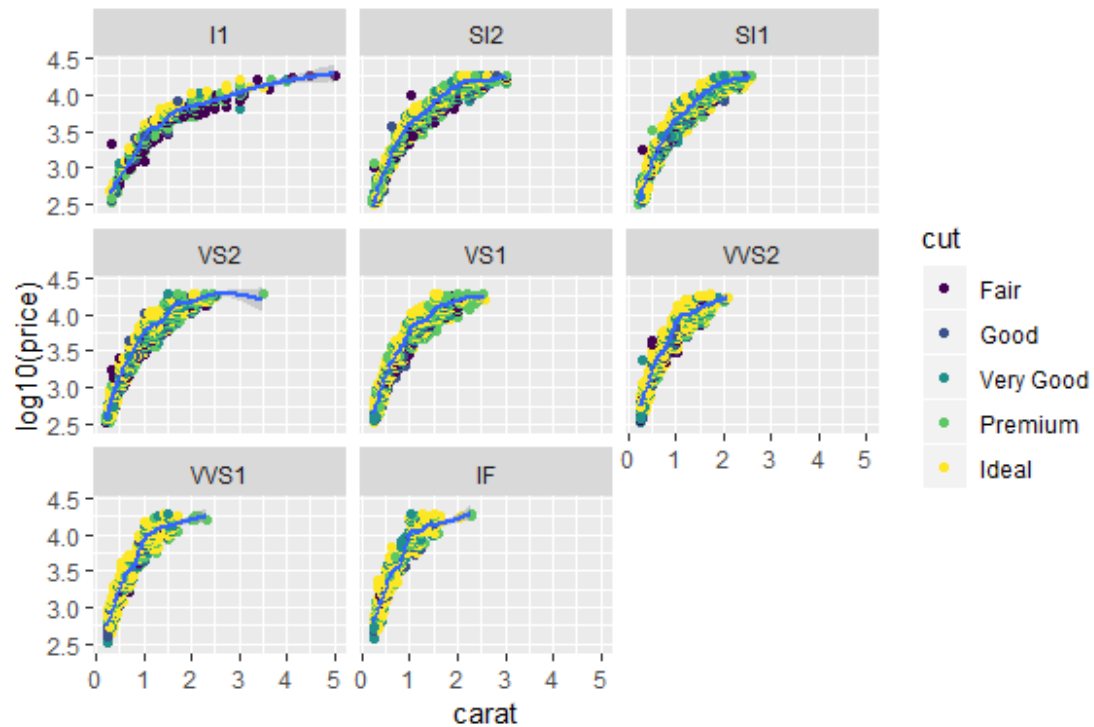
Change the scale to log10 scale.

```
ggplot(data = diamonds) + geom_point(mapping = aes(x = carat, y = log10(price)  
, color = cut)) + facet_wrap(~clarity)
```



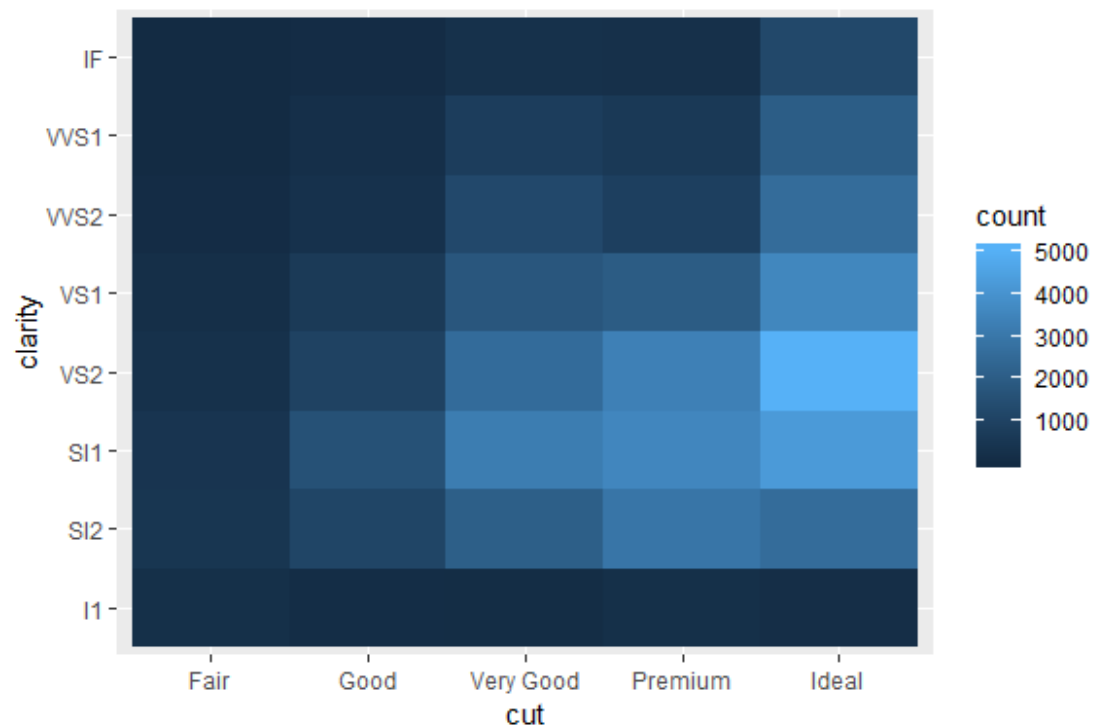
Add a smooth curve to the above scatter plot.

```
ggplot(data = diamonds) + geom_point(mapping = aes(x = carat, y = log10(price)  
, color = cut)) + facet_wrap(~clarity) + geom_smooth(mapping = aes(x = carat, y =  
log10(price)))
```



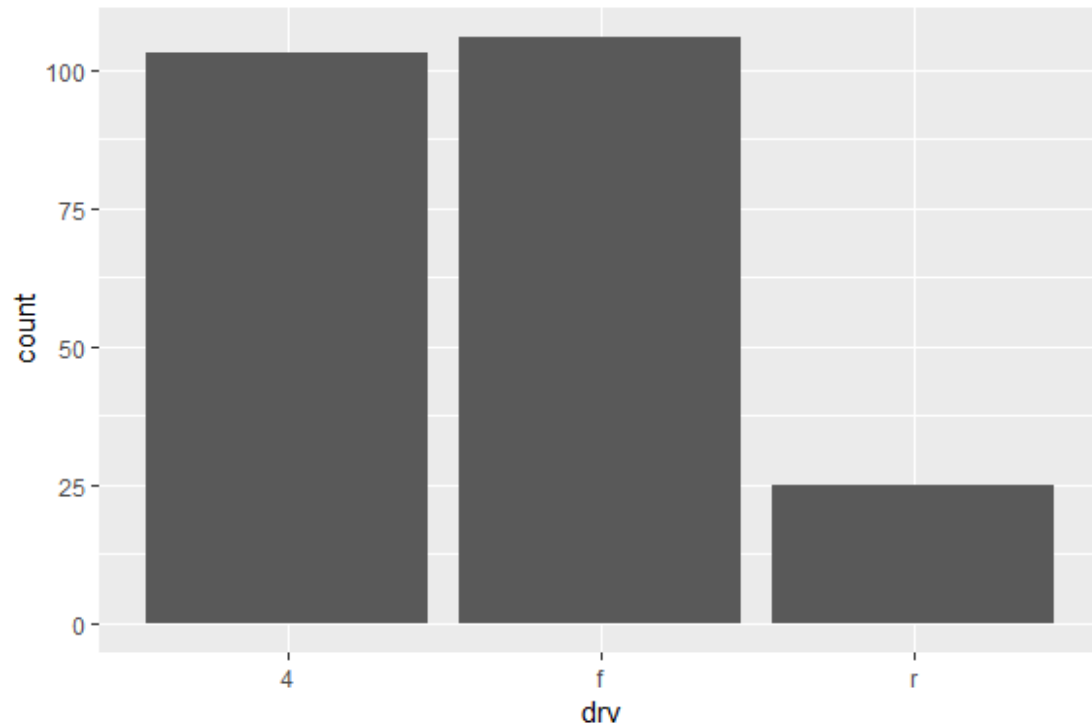
Create a tile plot (using `geom_bin2d()`) showing the relationship between cut and clarity.

```
ggplot() + geom_bin2d(mapping = aes(cut, clarity), data = diamonds)
```



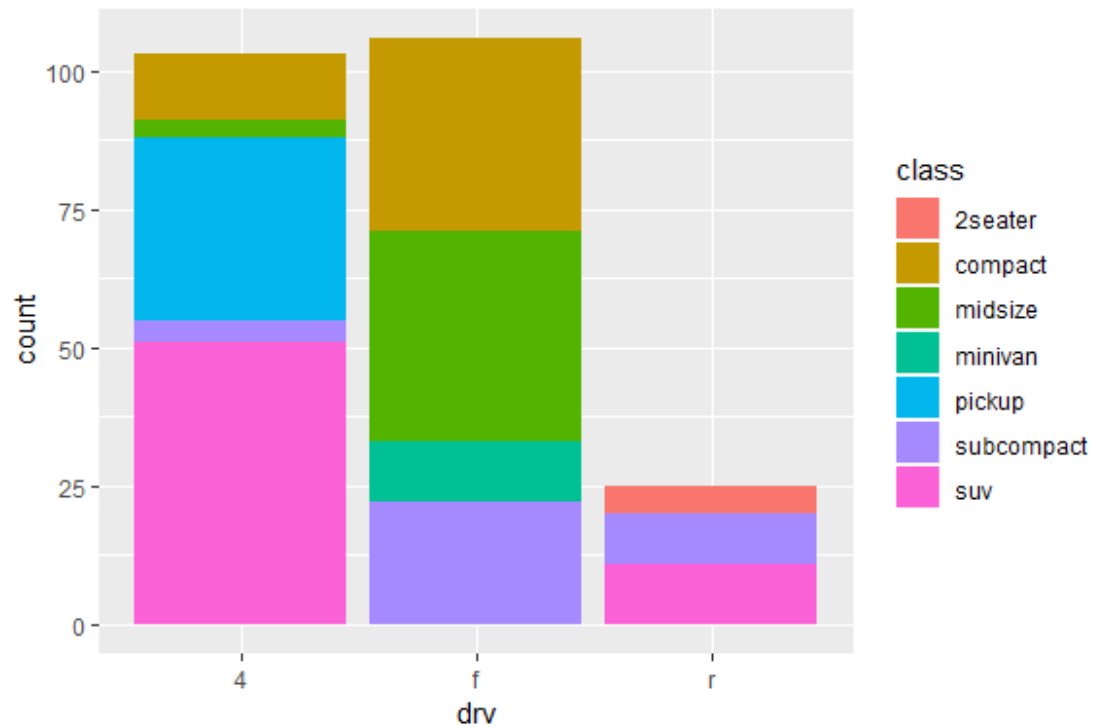
Using the mpg data set, create a bargraph showing the distribution of drv.

```
ggplot(data = mpg) + geom_bar(mapping = aes(drv))
```



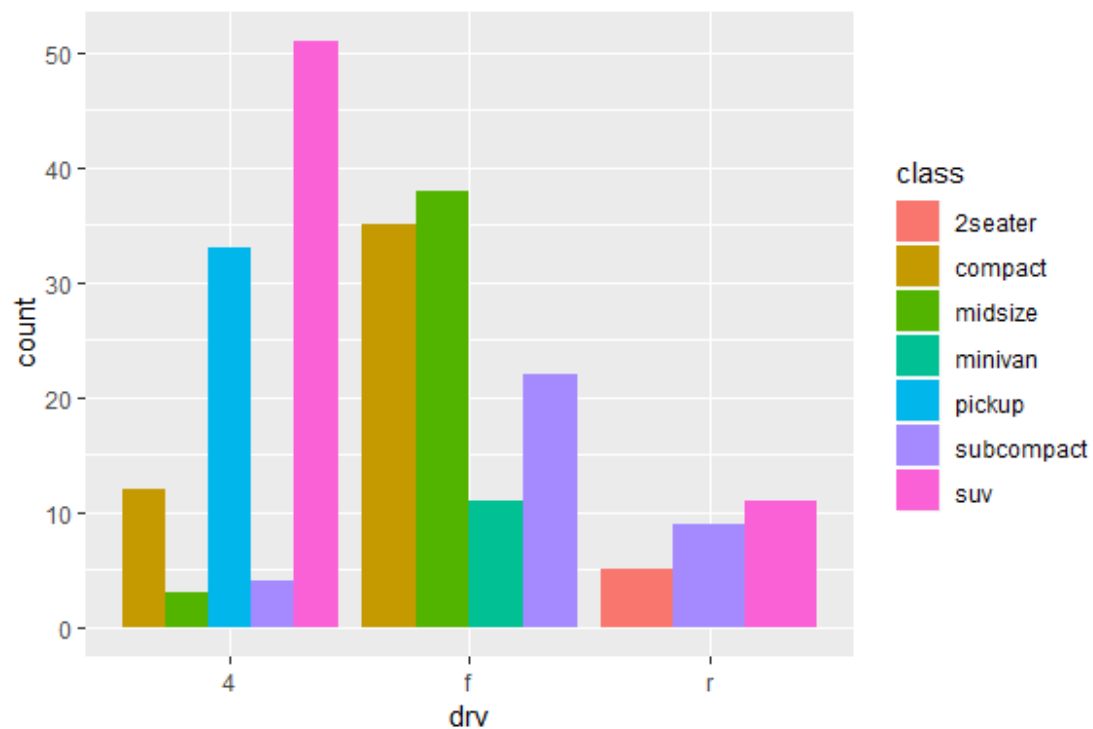
Add variable class to create a stacked bargraph (3 points)

```
ggplot(data = mpg) + geom_bar(mapping = aes(drv, fill = class))
```



Add class to the bargraph in a way that we can compare conditional probabilities.

```
ggplot(data = mpg) + geom_bar(mapping = aes(x = drv, fill = class), position = "dodge")
```



Based on the bargraph of the previous part, what is the most common class of car for 4-wheel drive cars?

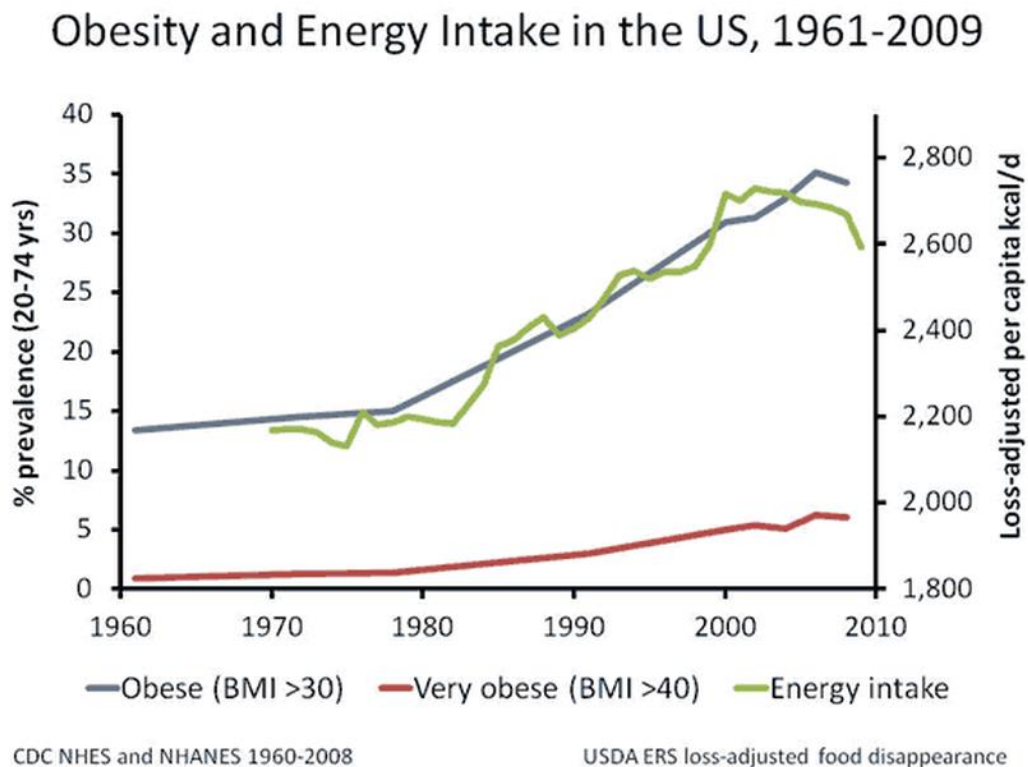
SUV

Using a side-by-side boxplot to compare the high way fuel efficiency (hwy) for different drives (drv). Which drive has the highest high way fuel efficiency?

Full-wheel drive

Dissecting graphs

The chart that might explain the obesity crisis In this graph, Stephan J. Guyenet, an obesity researcher at the University of Washington, plots the rise of obesity (in red and blue) and the increase in food consumption (in green) over the past few decades. Since 1970, the incidence of obesity has more than doubled to about 35 percent of adults. At the same time, Americans are eating roughly 400 more calories a day, an increase of about 20 percent. Guyenet argues that the mathematics works out such that this one trend - the increase in calorie consumption - can account for the entire rise in obesity.



Dissect the graph to the four elements: visual cues, coordinate system, scale, and context.

- Visual cues

The author uses a time series directional plot. The author also qualitative colors that represent specific groups of people and how much they are eating.

- Coordinate system

Cartesian coordinate system

- Scale

The author uses a linear time scale for the x axis and for the other vertical axes, the author uses a linear numeric scale.

- Context

The author uses the x axis to show the specific year. The author uses the y axis to to show the percentage of people within the ages of 20 and 24 that are obese. It also uses the vertical axis on the right to show the average number of calories being consumed. The author gives a title to the entire data visualization showing the correlation between energy intake and obesity rates between the years of 1961-2009. The author gives a subtitle to show who collected the data and from what time periods.