1. Use the built-in dataset mtcars for this assignment. Do cars with big engines use more fuel than cars with small engines? You probably already have an answer,

but try to make your answer precise. What does the relationship between engine size and fuel efficiency look like? Is it positive?

Negative?

source code:

```
cor(mtcars$wt, mtcars$mpg)

data(mtcars)

plot(mtcars$disp, mtcars$mpg, xlab = "Engine size", ylab = "Miles per gallon", main = "Scatter plot of engine size vs mpg")
```

output :

```
> summary(mtcars$mpg)

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.

 10.40   15.43   19.20   20.09   22.80   33.90


> summary(mtcars$disp)

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.

 71.10  120.80  196.30  230.72  326.00  472.00


> cor(mtcars$mpg, mtcars$disp)

[1] -0.8475514
```

2.How many rows are in mpg? How many columns?

sorce code:

```
data(mpg)


dim(mpg)
```

output :

[1] 234 11

3.Which variables in mpg are categorical? Which variables are continuous?

source code:

data(mpg)

str(mpg)

output:

```
> data(mpg)
> str(mpg)
'data.frame':    234 obs. of  11 variables:
 $ manufacturer: Factor w/ 15 levels "audi","chevrolet",..: 1 1 2 2 1 1 2 1 2 2 ...
 $ model       : Factor w/ 38 levels "4runner 4wd",..: 2 3 5 4 1 6 7 8 9 10 ...
 $ displ       : num  1.8 1.8 2 2 2.8 2.3 1.8 1.8 2 2.8 ...
 $ year        : int  1999 1999 2008 2008 1999 1999 2008 1999 2008 2008 ...
 $ cyl         : factor  4 4 4 4 6 4 4 4 4 4 ...
 $ trans       : factor  auto auto manual manual auto auto manual auto manual ...
 $ drv         : factor  f f f f r f r f 4 4 ...
 $ cty         : int  21 21 23 24 21 16 18 18 16 20 ...
 $ hwy         : int  29 29 31 30 26 27 26 25 28 27 ...
 $ fl          : factor  r r p p r r p r e d ...
 $ class       : factor  compact compact compact compact compact compact compact compact compact compact ...
```

4.Take the first faceted plot in this section:

ggplot(data = mpg) +

geom_point(mapping = aes(x = displ, y = hwy)) +

facet_wrap(~ class, nrow = 2)

What are the advantages to using faceting instead of the colour aesthetic?

1.Faceting allows us to compare the relationship between two variables across multiple subsets of the data, while color can only distinguish between a limited number of categories.

2.Faceting can be used to create a larger number of panels than the number of colors that can be easily distinguished, allowing for more detailed comparisons.

3.Faceting can be more accessible for colorblind viewers who may have difficulty distinguishing between certain colors.

What are the disadvantages? How might the balance change if you had a larger dataset?

Disadvantages of using faceting:

Faceting can make the plot more cluttered and difficult to interpret, especially if there are many panels.

Faceting requires more space to display the plot, which may be an issue in situations where space is limited.

source code:

```
library(ggplot2)

data(mpg)

ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_wrap(~ class, nrow = 2)
```

5.What geom would you use to draw a line chart? A boxplot? A histogram? An area chart?

The geom to use for each type of chart are:

Line chart: geom_line()

Boxplot: geom_boxplot()

Histogram: geom_histogram()

Area chart: geom_area()

geom_line(): This geom is used to create line charts, which show trends or changes in a variable over time or across different groups.

geom_boxplot(): This geom is used to create boxplots, which display the distribution of a variable across different categories or groups.

geom_histogram(): This geom is used to create histograms, which display the distribution of a variable as a series of vertical bars.

geom_area(): This geom is used to create area charts, which are similar to line charts but with the area beneath the line filled in to emphasize the magnitude of the values.