

Tutorial Exercises Week 5

Question 1

The S&P 500 is a stock market index that tracks the stock market performance of the 500 biggest publicly-listed companies in the US. The file [SP500.csv](#) contains the value of this index for each day from 2015 until 2023.

With `ggplot`, if you want to plot a variable `x` over time `t` from the dataframe `df`, you can use the command: `ggplot(df, aes(t, x)) + geom_line()`.

Use this approach to plot the closing price (variable `close`) over time.

Based on your plot, during which of the following periods did the index experience the largest crash?

- At the beginning of 2016.
- At the end of 2018.
- Near the beginning of 2020.
- In the second half of 2023.

Question 2

Using the S&P 500 data, create a variable in your dataset which is the percentage change in the closing price from one day to the next. Call this the daily return.

If x_t is the closing price on date t , then the percentage change in the closing price from the previous day is given by the following equation:

$$100 \times \frac{x_t - x_{t-1}}{x_{t-1}}$$

Plot the daily returns over time using `ggplot` with the `geom_line()` function.

Which of the following is true in the plot?

- Early 2020 had both the biggest positive and negative daily returns.
- Early 2020 had the biggest positive daily returns, but not the biggest negative daily returns.
- Early 2020 had the biggest negative daily returns, but not the biggest positive daily returns.

- Early 2020 had neither the biggest positive daily returns, nor the biggest negative daily returns.

Question 3

Download the dataset [toyota-camry-ads.csv](#). This contains information on classified advertisements placed on [Craigslist](#) for used Toyota Camry cars, one of the most common sedan cars in the US. The variables are:

- **condition**: what kind of condition the car is in (good, like new, etc).
- **odometer**: how many miles the car has travelled in its lifetime (the car's "mileage").
- **paint_color**: the car's color.
- **price**: the asking price of the car.
- **year**: the year the car was bought new.

Create a histogram of the **year** variable. Based on this histogram, choose the answer below which contains a correct statement about the data.

- The majority of cars in the dataset were bought after 2000.
- The majority of cars in the dataset were bought between 2007-2009.
- The majority of cars in the dataset were bought after 2010.
- All cars in the dataset were bought after 1990.

Question 4

Create a bar plot of the **condition** variable. Based on this plot, what is the most common condition for Toyota Camry cars for sale on this site to be in?

Plotting tip: `ggplot` will by default order the "conditions" alphabetically. But sometimes it makes more sense for categorical variables like this to be ordered differently. We might want to order the bars from worst condition to best condition. We can do this by converting the condition variable to a factor variable and specifying the order we want:

```
ordered_levels <- c("salvage", "fair", "good", "excellent", "like new", "new")
df$condition <- factor(df$condition, levels = ordered_levels)
```

Variables like this are called "ordinal".

Question 5

Create a scatter plot with:

- **odometer** on the horizontal axis.
- **price** on the vertical axis.

In addition, make the colors of the points represent the values of the **year** variable.

Choose the answer below which best interprets this scatter plot.

Hint: To make a scatter plot with `x` and `y` with colors representing `z` we do `ggplot(df, aes(x, y, color = z)) + geom_point()`

- Higher values of odometer are usually associated with lower prices. In addition, newer cars usually sell for a higher price.
- Higher values of odometer are usually associated with lower prices. In addition, older cars usually sell for a higher price.
- Higher values of odometer are usually associated with higher prices. In addition, newer cars usually sell for a higher price.
- Higher values of odometer are usually associated with higher prices. In addition, older cars usually sell for a higher price.