1. Read in the pumpkins.csv data.

2. Identify the heaviest pumpkin grown in any of the competitions – what variety was it? Where was it from and when was it grown?

Heaviest Pumpkin: Baby Bear grown in Chicago, California, Mexico during 2014-P

3. Write a function to change the weight in pounds (lbs) to kilograms (kg) and use this function to create a new column in your data set called weight\_kg.

4. Create another new column in your data set called weight\_class. This column should be a factor with three levels: light, medium and heavy and should be based on the weight of the pumpkins. You can choose the thresholds for these three groups yourself.

5. Plot the relationship between the estimated weight and actual weight of the pumpkins. This can be in either lbs or kg but remember to ensure that both axes are in the same units. Colour the points in your plot based on the weight\_class column that you created in question 4. Ensure the plot has appropriate axis labels and is clear and well presented. Save this plot to your computer.图表, 散点图

描述已自动生成

This plot shows the relationship between the estimated and actual weights of pumpkins. Different colors indicate the weight classes (light, medium, heavy). There is a general linear trend, but some deviations can be observed, particularly in the medium class.

6. Filter the data to include only pumpkins from three countries of your choosing and save this filtered data set to your computer in csv format.

7. Summarise your filtered dataset from question 6:a. Identify the mean weight of pumpkins for each of your three countries. Which country had the highest mean pumpkin weight?

The Highest mean pumpkin weight: Canada

b. Identify the mean weight for each variety of pumpkin for each of your countries. Which variety in which country had the lowest mean weight?

Lowest mean variety and country:

country USA

variety Baby Bear

weight\_lbs 726.52564

Name: 4, dtype: object

8. Using your filtered data set from question 6, plot pumpkin weight distributions (in either lbs or kg) for your three countries as a boxplot. Ensure the plot has appropriate axis labels and is clear and well presented. Save this boxplot to your computer.

图表, 箱线图

描述已自动生成

This boxplot illustrates the distribution of pumpkin weights across three selected countries. Canada shows the highest median weight, while other countries exhibit a broader range of weights.

9. Redraw your plot from question 8 as a facet plot showing the data from each variety of pumpkin as a separate sub-plot. Save this plot to your computer.图表, 箱线图

描述已自动生成

This facet plot divides the weight distributions by pumpkin variety and country. It highlights the variability of each pumpkin variety in different regions, with some varieties (e.g., Baby Bear) showing significant differences between countries.