The Weighted Mean and the Median: Takeaways

Syntax

- Learning namespace:

```
name_of_package::name_of_function(parameters)
```

- Computing the weighted mean of any numerical vector using R base function weighted.mean():

```
mean  <- weighted.mean(x = distribution, w = weights)
```

- Creating our weighted mean function:

```
compute_weighted_mean <- function(distribution, weights) {
    weighted_distribution  <-  distribution * weights
    sum(weighted_distribution) / sum(weights)
}
```

- Finding the median of any numerical vector using the function median():

```
median  <- median(vector)
```

- Generating a boxplot for a single column to confirm if it has outliers:

```
library(ggplot2)
ggplot(data = df,
    aes(x = "", y = column_name)) +
    geom_boxplot() +
    xlab("column name") +
    ylab("")
```

Concepts

- When data points bear different weights, we need to compute **the weighted mean**. The formulas for the weighted mean are the same for both samples and populations, with slight differences in notation:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i} = \frac{x_1 w_1 + x_2 w_2 + \ldots + x_n w_n}{w_1 + w_2 + \ldots + w_n}$$

$$\mu = \frac{\sum_{i=1}^{N} x_i w_i}{\sum_{i=1}^{N} w_i} = \frac{x_1 w_1 + x_2 w_2 + \ldots + x_n w_N}{w_1 + w_2 + \ldots + w_N}$$

- It's difficult to define the median algebraically. To compute the median of a vector, we need to:
    - Sort the values in an ascending order.
    - Select the middle value as the median. If the distribution is even-numbered, we select the middle two values, and then compute their mean — the result is the median.
- The median is ideal for:
    - Summarizing numerical distributions that have **outliers**.
    - **Open-ended** distributions.
    - **Ordinal data**.

Resources

- An intuitive introduction to the weighted mean.
- The Wikipedia entry on the weighted mean.
- The Wikipedia entry on the median.
- Useful documentation:
    - weighted.mean() function
    - median() function