

# STA130 T0209

Week 3

(Materials used in this presentation are provided by the U of T Statistical Sciences Department.

This presentation was prepared by Vivian Ngo.)

# Reminders

- Homework: incomplete or late = 0
  - Corrupted files
- Email policy
- Office hours, TA hours, Piazza
- Tutorial: 10 minutes after the hour
  - Missing tutorials
- Writing assignments
  - Include graphs
  - Title your graphs
  - Concise
  - Value added

Instructor/TA	When	Where
Dr. White	Mondays 12 (after class) - 2pm	SS 6006
Dr. Moon	Mondays 4 (after class) - 5:30pm	SS 6024A
TA (Various)	Tuesdays 11am - 2pm (Starting Jan 15th)	SS 623B (basement of SS)
TA (Various)	Wednesdays 11am - 2pm (Starting Jan 9th)	SS 623B (basement of SS)
TA (Various)	Thursdays 11am - 2pm (Starting Jan 10th)	SS 623B (basement of SS)

# Agenda

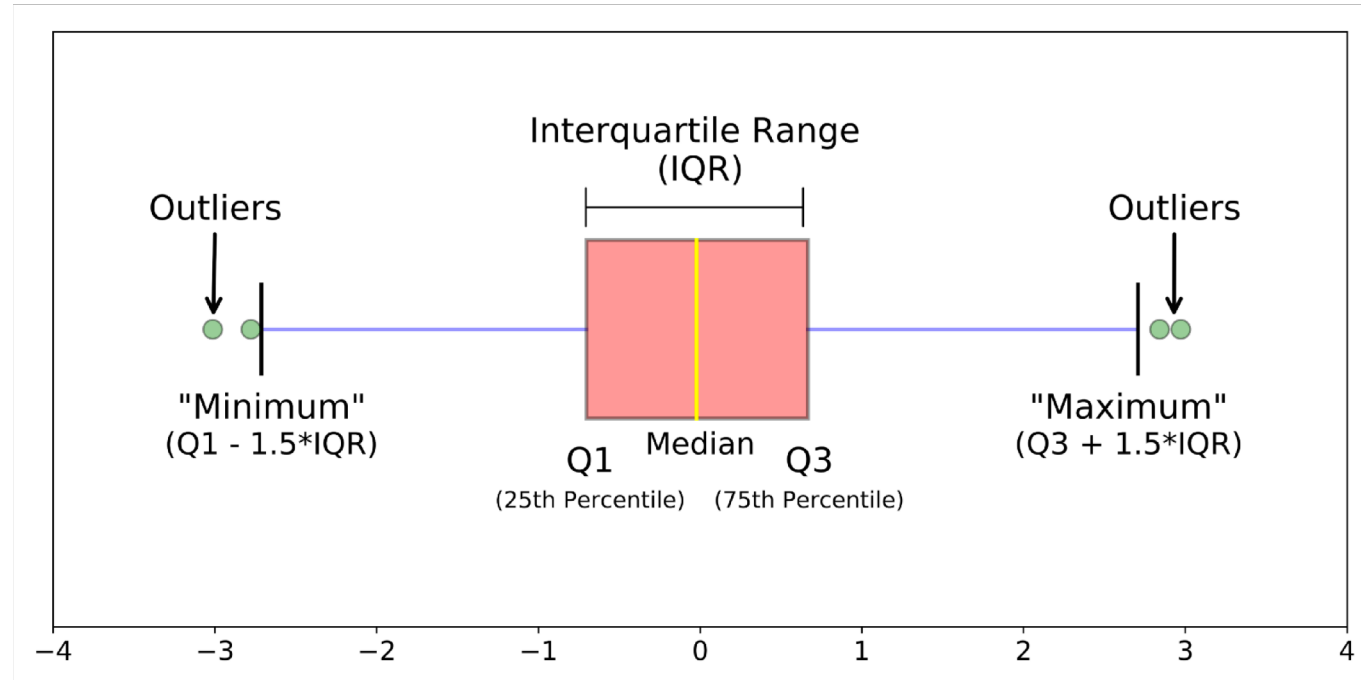
- Vocabulary
- Boxplots
- Homework discussion
- Writing activity
- Second half: Meeting mentors

# Vocabulary

- General terms:
  - Outlier
  - Interquartile range
  - Boxplot
  - Proportion
- Terms for describing data wrangling:
  - Cleaning data
  - Tidy data
  - Create a dummy variable
  - Removing the column
  - Replace values above/below a certain threshold
  - Taking the subset of variables
  - Filtering the data frame based on a condition (e.g. based on one of the variables/columns)
  - Renaming the variables
  - Grouping the categories

# Boxplots

- When to use?
- What information do they display?



# Homework discussion

- For Question 1D, you used both histograms and boxplots to visualize your data.
  - Which features were easier/harder to observe from each of the visualizations?
  - In what situations may you want to choose a boxplot over a histogram, or vice versa? Explain.

# Writing activity

- Self-Reflection:
  - What questions, if any, do you have so far regarding the course materials?
  - What is one of your favorite things about tutorial? Least favorite?