

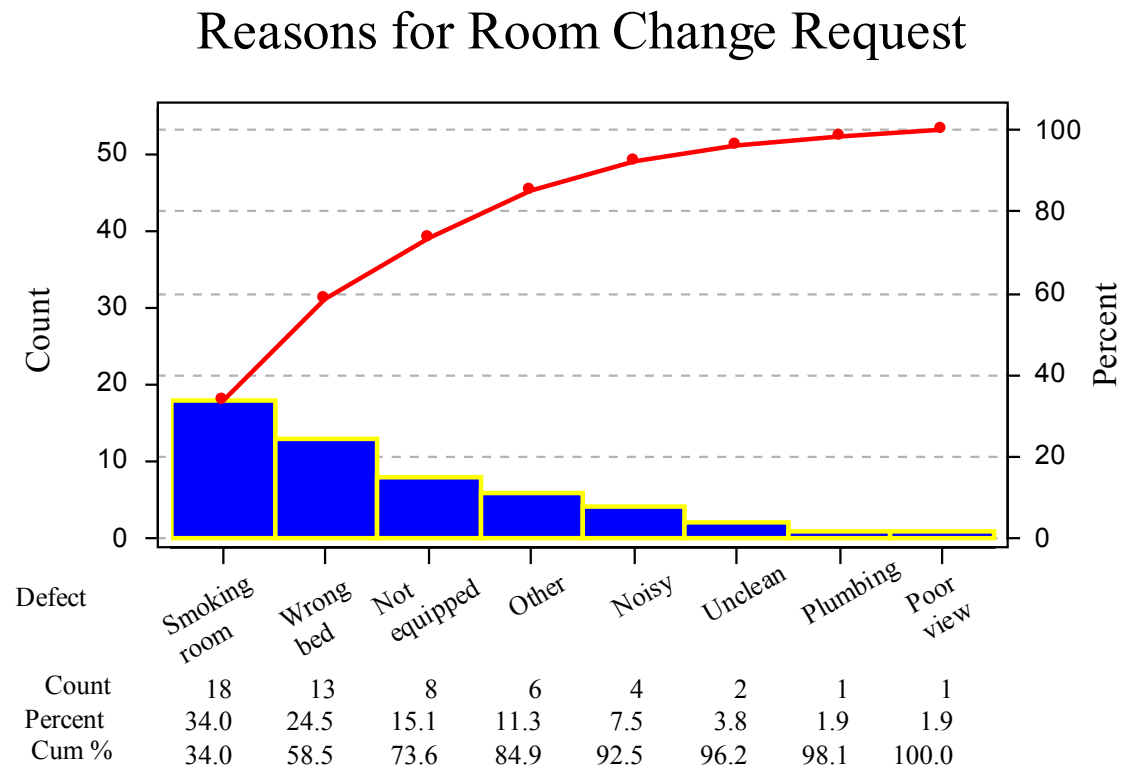
3. GRAPHICAL SUMMARIES OF DATA

- A picture is worth a thousand words.
- Using graphical displays, we can make statistical data come alive.
- A well-presented picture (chart) can reveal unexpected patterns.

For categorical data, we can use a **Pareto Chart**.

The data for 53 hotel guests are classified according to the primary reason why the guests requested new rooms. A pareto chart gives a bar chart for the counts, with categories ordered from most frequent to least frequent. A plot of cumulative percentages is superimposed.

We see, for example, that “Smoking room” and “Wrong bed” are the two categories that most frequently occur, accounting for over 58 percent of the total.



For a batch of numerical (quantitative) data, we can use a **histogram**. The data are divided into classes, usually equally spaced intervals, and vertical bars are drawn for each interval. The height of the bar is the frequency, proportion or percentage of data in the given interval.

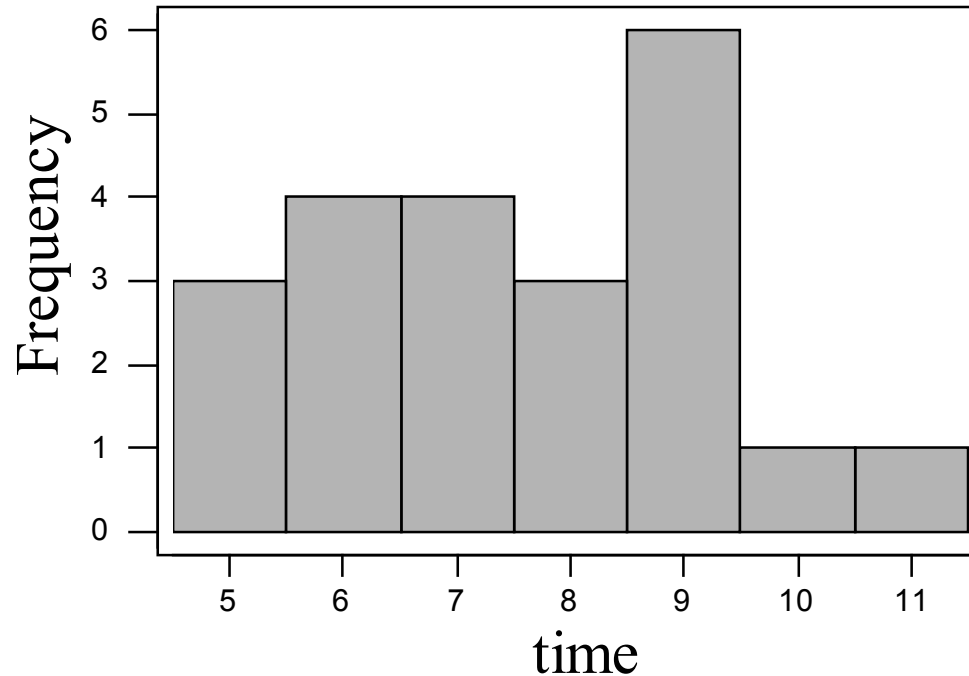
Frequency is the number of times data falls in an interval.

Proportion, (also known as relative frequency) is frequency divided by total number of data points.

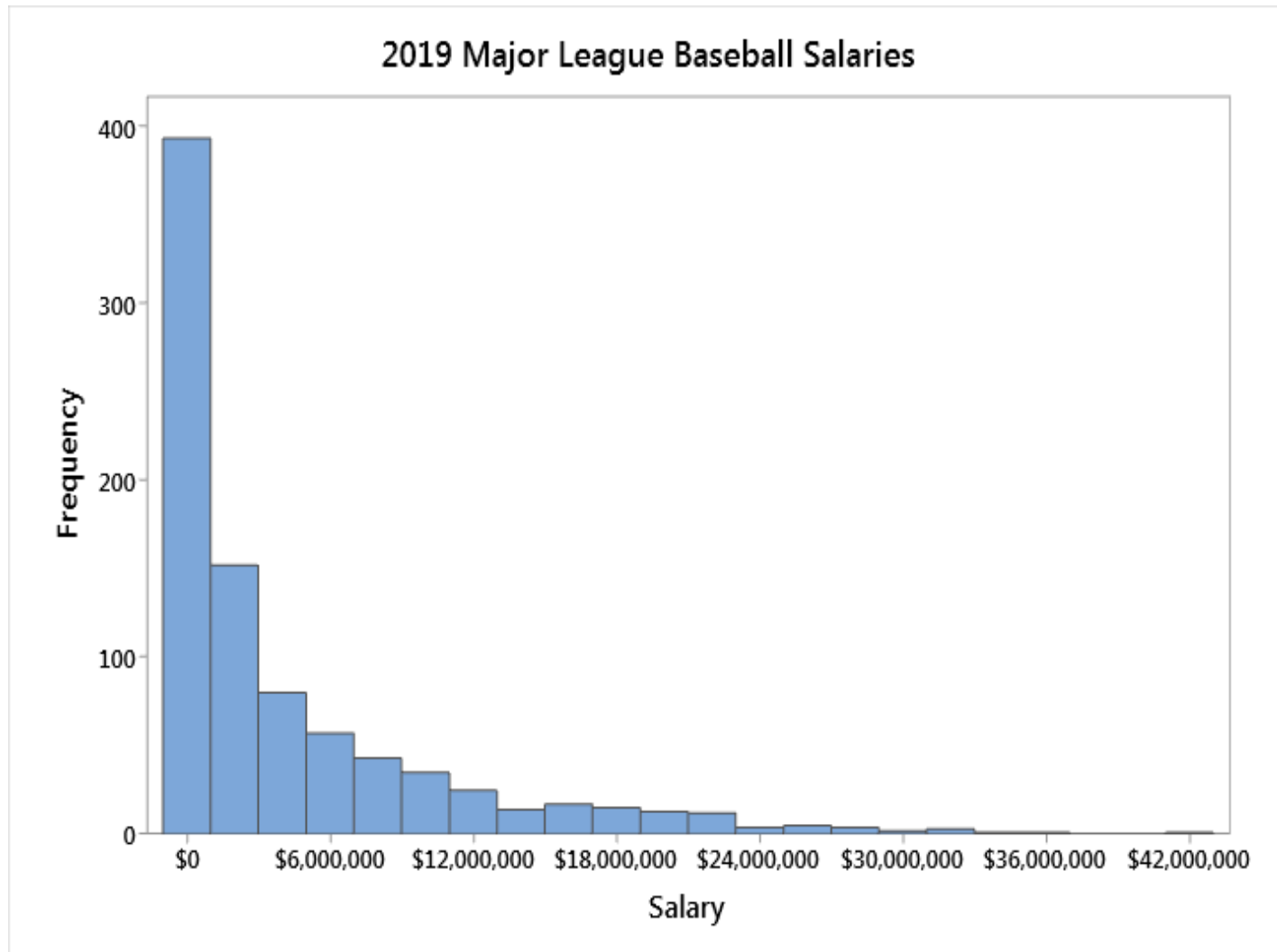
Percentage is proportion times 100.

Eg: Time in seconds for 0 to 60 mph for 22 German sports cars.

10.0, 7.9, 7.1, 8.6, 6.4, 6.9, 8.7, 8.3, 8.5, 6.4, 7.5,
6.7, 5.5, 6.0, 5.4, 6.9, 5.1, 4.9, 8.5, 8.8 10.9, 8.9.



Eg: Histogram of Annual Salaries of Major League Baseball players for 2019.

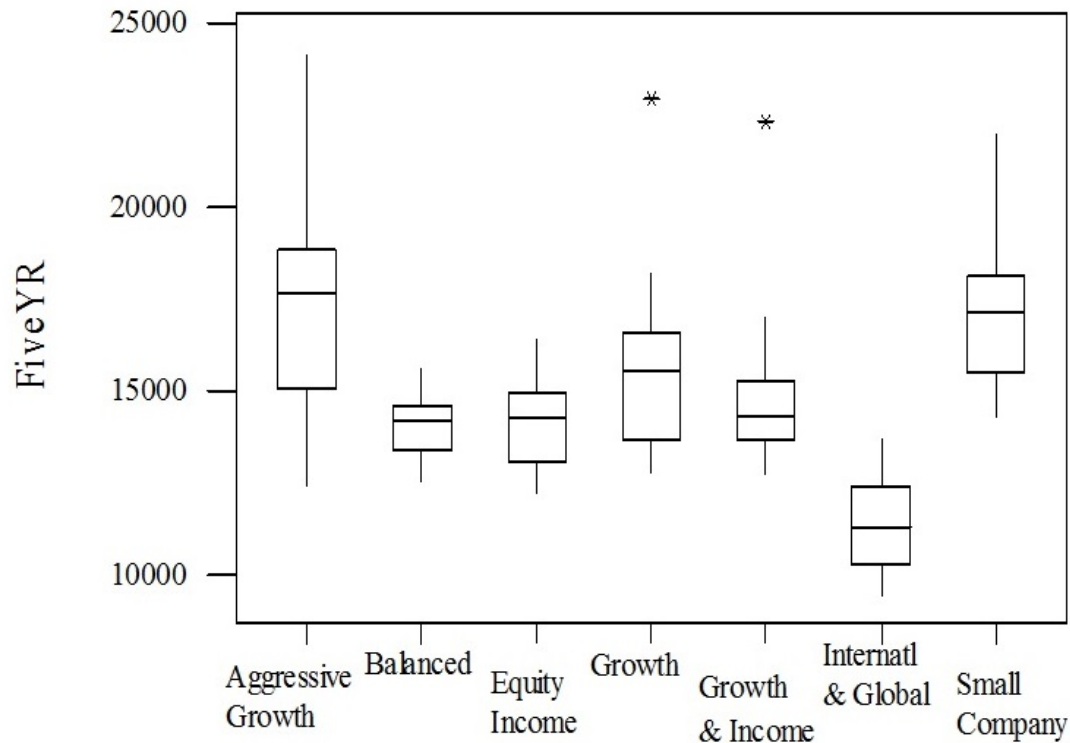


Boxplots

For numerical data. Summarizes basic distributional properties. (center, spread, skewness, outliers). Useful for comparing two or more batches of data, side-by-side. We will give details later.

Five Year Performance of Mutual Funds

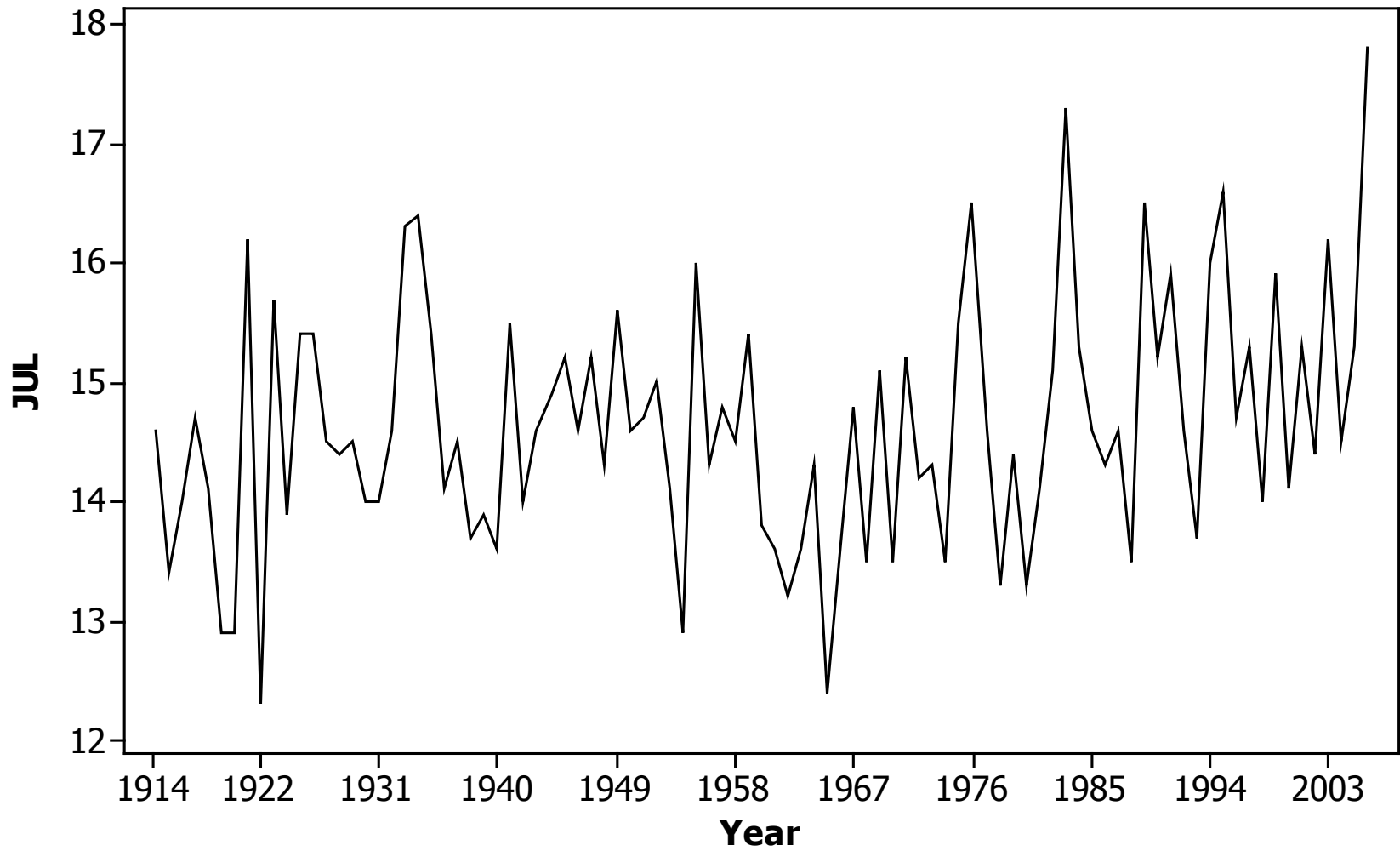
Based on a \$2000/yr investment



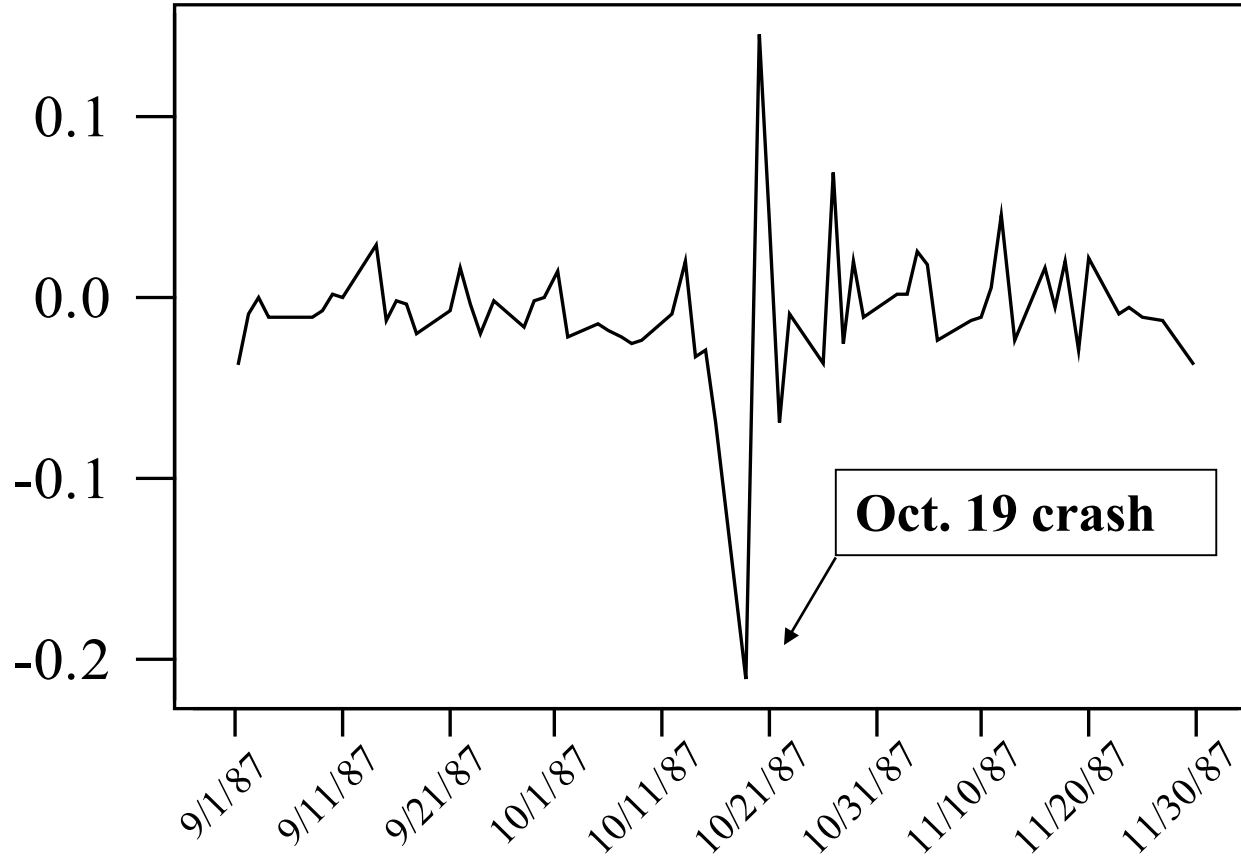
For data collected over time, we can use a **Time Series Plot**.



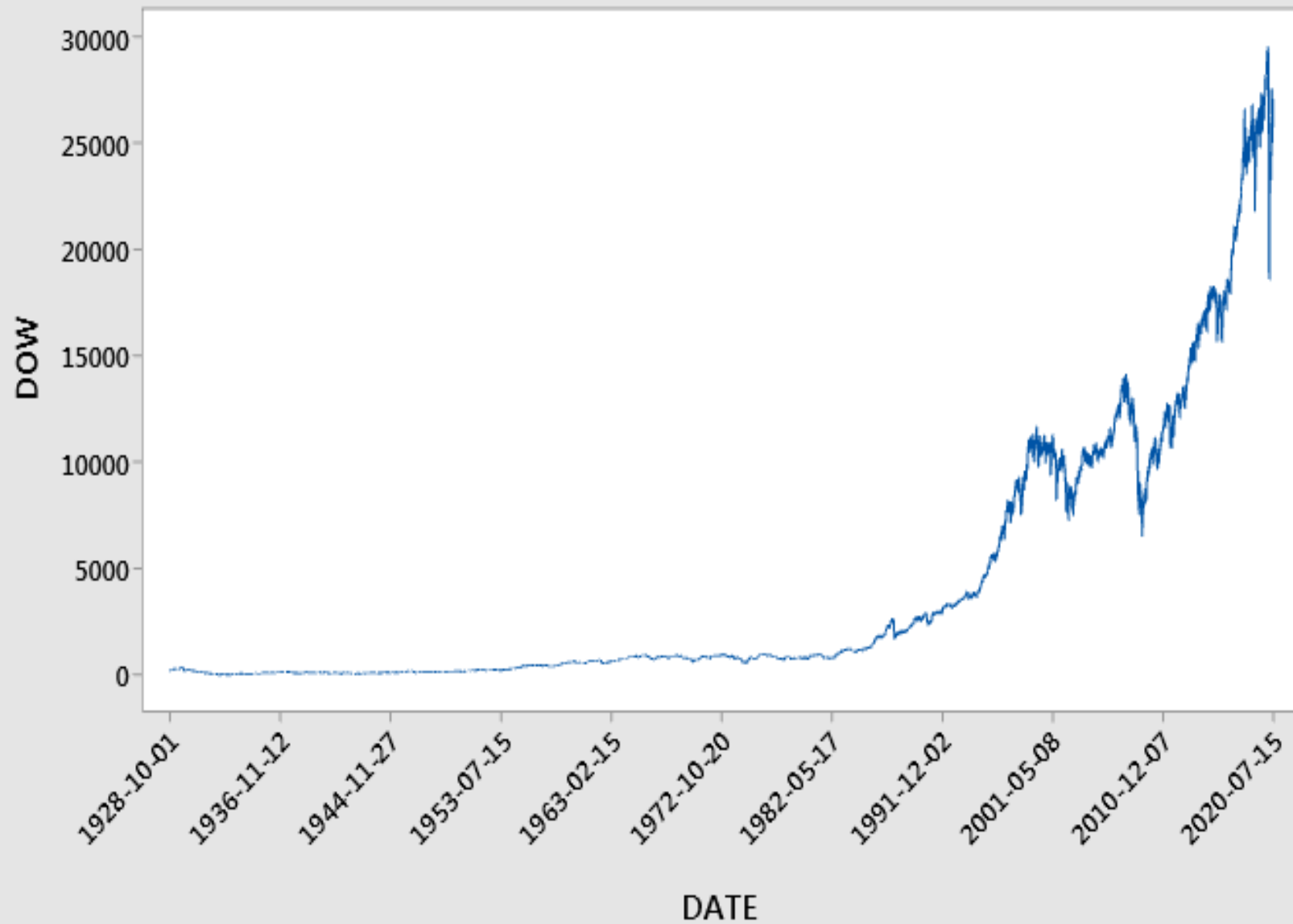
UK July Mean Temperatures (Degrees C)



GM Returns



Time Series Plot of DOW



Scatterplots

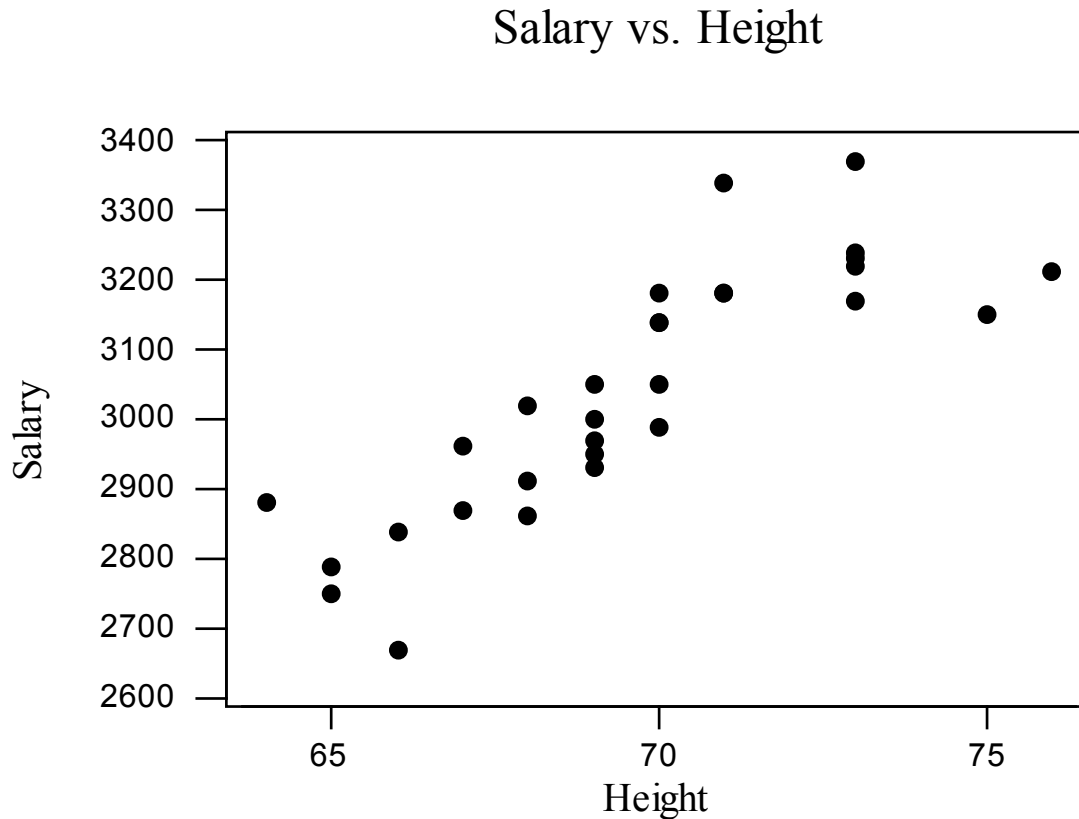
Suppose we measure two variables, X and Y .

A **scatterplot** is a two-dimensional plot of the Y value (on the vertical axis) versus the corresponding X value (on the horizontal axis).

We call this a plot of Y versus X .

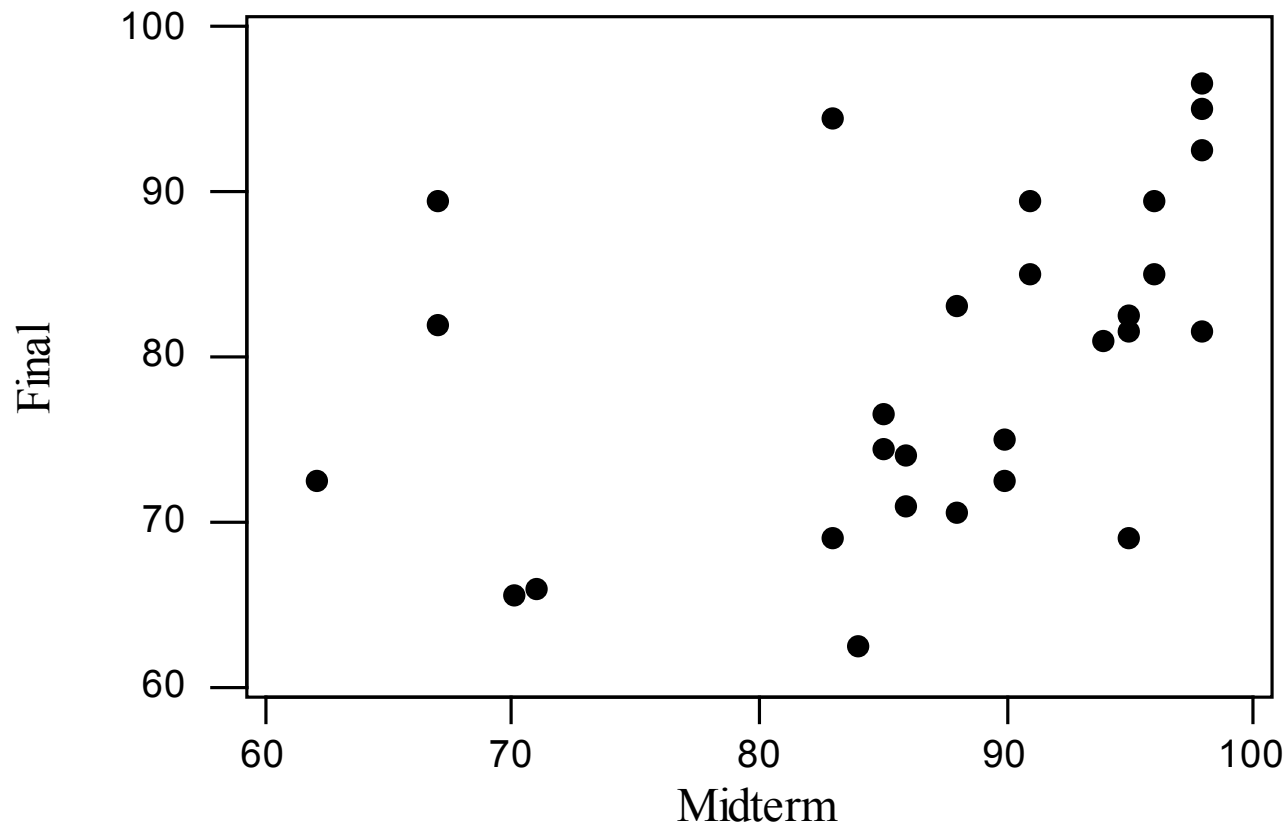
- Y goes FIRST!!

Eg: Monthly salary for the first post-MBA job vs. heights of 1986 male MBA graduates of the University of Pittsburgh (from the Wall Street Journal).



Eg: Scores on Final vs. Scores on Midterm of students from a recent MBA Statistics course.

Final Exam vs. Midterm Grades



Eg: Scatterplot for Old Faithful.

Time Between Eruptions vs. Eruption Duration

