Lecture 6

Descriptive Statistics IV

Text: Chapter 3

STAT 8010 Statistical Methods I
September 2, 2019

Whitney Huang
Clemson University

Agenda

1. Percentiles and Quartiles
2. Boxplots
3. Z-scores & Empirical Rule
4. Visualizing Time Series, Cross-Sectional, and Spatio-Temporal Data sets

Percentiles

- The p_{th} percentile is a value such that at least $p\%$ of the data set is less than or equal to this value

- Calculation of percentiles using the indexing method:
  1. Sort the set of numbers in an increasing order
  2. For the p_{th} percentile, compute the index $i = \frac{np}{100}$ where n is the sample size
  3. If i is an integer then p_{th} percentile is the average of i_{th} value and $(i+1)_{\text{th}}$ value, otherwise take the $(i+1)_{\text{th}}$ value

- Quartiles:
  1. $Q1$: first quartile
  2. M $(Q2)$: median (second quartile)
  3. $Q3$: third quartile
  4. Interquartile range or IQR: $Q3 - Q1$

Example

Find Q_1, M, Q_3 and IQR of the following list of values: 13, 18, 13, 14, 13, 16, 14, 21, 13 using the indexing method

1. Order the data first: 13, 13, 13, 13, 14, 14, 16, 18, 21
2. Find the sample size n and compute the indices for $p = 25, 50, 75$
3. $n = 9 \Rightarrow$ the indices are $3, 5, 7 \Rightarrow Q_1 = 13$, $M = 14$, $Q_3 = 16$
4. IQR $= Q_3 - Q_1 = 16 - 13 = 3$

Steps to Making a Boxplot

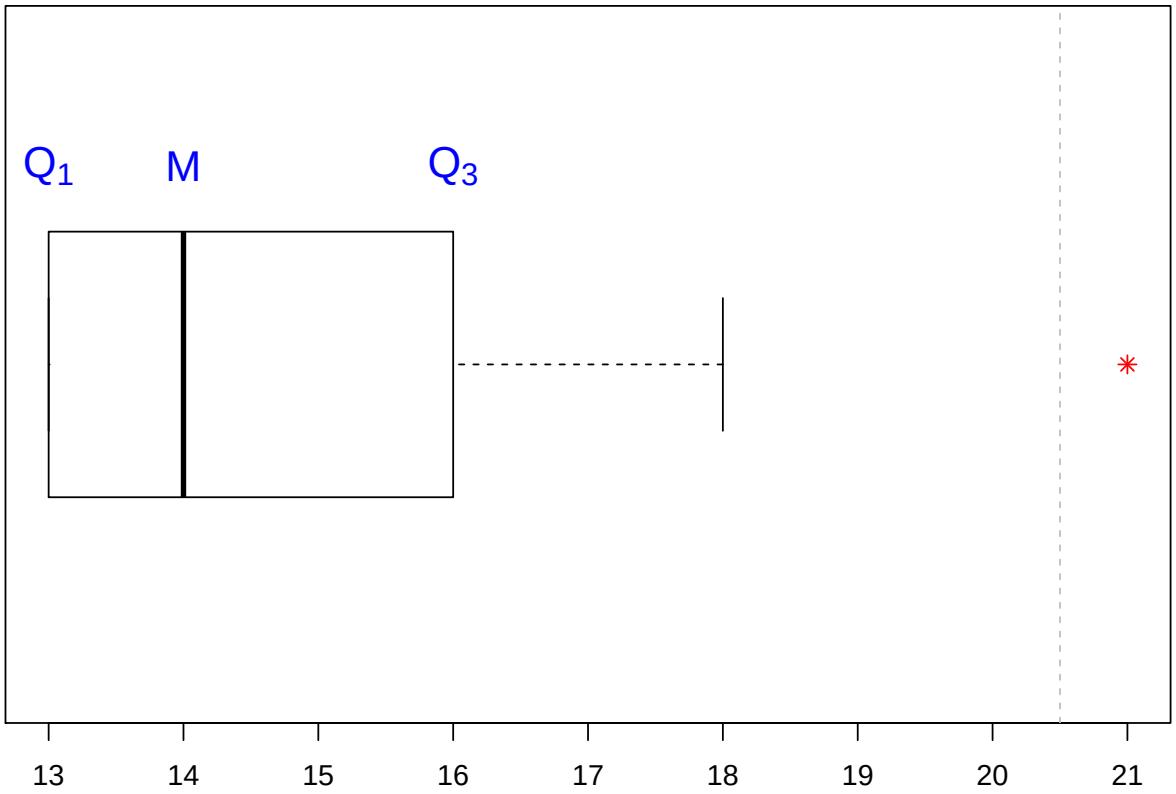
1. Find Q_1, M, Q_3 and draw a box from Q_1 to Q_3. Add a vertical line inside the box at M

2. Compute the value of Lower Fence (LF) $= Q1 - 1.5$IQR and the Upper Fence (UF) $= Q3 + 1.5$IQR. Find the largest value $\leq$ UF and the smallest value $\geq$ LF. Draw whiskers go from Q_1, Q_3 to these two values

3. Plot the individual outlier(s) (i.e., the values either $>$ UF or $<$ LF)

Bopxplot

- **Ordered data values:** 13, 13, 13, 13, 14, 14, 16, 18, 21
- **IQR** $16 - 13 = 3 \Rightarrow$ LF = $13 - 1.5 \times 3 = 8.5$; UF = $16 + 1.5 \times 3 = 20.5$



Example

Suppose we have the following list of values: 13, 18, 13, 14, 13, 16, 14, 21, 13, 9, 27, 18, 25, 20, 6

- Find the 35th percentile

Example

Suppose we have the following list of values: 13, 18, 13, 14, 13, 16, 14, 21, 13, 9, 27, 18, 25, 20, 6

- Find the 35th percentile
  1. Sort the data:
     $6, 9, 13, 13, 13, 13, 14, 14, 16, 18, 18, 20, 21, 25, 27$

Example

Suppose we have the following list of values: 13, 18, 13, 14, 13, 16, 14, 21, 13, 9, 27, 18, 25, 20, 6

- Find the 35th percentile
  1. Sort the data:
     $6, 9, 13, 13, 13, 13, 14, 14, 16, 18, 18, 20, 21, 25, 27$
  2. Compute the index value $i = \frac{35 \times 15}{100} = 5.25 \Rightarrow$ the 35th percentile is 13

Example

Suppose we have the following list of values: 13, 18, 13, 14, 13, 16, 14, 21, 13, 9, 27, 18, 25, 20, 6

- Find the 35th percentile
  1. Sort the data:
     $6, 9, 13, 13, 13, 13, 14, 14, 16, 18, 18, 20, 21, 25, 27$
  2. Compute the index value $i = \frac{35 \times 15}{100} = 5.25 \Rightarrow$ the 35th percentile is 13

Example

Suppose we have the following list of values: 13, 18, 13, 14, 13, 16, 14, 21, 13, 9, 27, 18, 25, 20, 6

- Find the 35th percentile
  1. Sort the data:
     $6, 9, 13, 13, 13, 13, 14, 14, 16, 18, 18, 20, 21, 25, 27$
  2. Compute the index value $i = \frac{35 \times 15}{100} = 5.25 \Rightarrow$ the 35th percentile is 13

- Find the 65th percentile

Example

Suppose we have the following list of values: 13, 18, 13, 14, 13, 16, 14, 21, 13, 9, 27, 18, 25, 20, 6

- Find the 35th percentile
  1. Sort the data:
     $6, 9, 13, 13, 13, 13, 14, 14, 16, 18, 18, 20, 21, 25, 27$
  2. Compute the index value $i = \frac{35 \times 15}{100} = 5.25 \Rightarrow$ the 35th percentile is 13

- Find the 65th percentile
  1. Sort the data:
     $6, 9, 13, 13, 13, 13, 14, 14, 16, 18, 18, 20, 21, 25, 27$

Example

Suppose we have the following list of values: 13, 18, 13, 14, 13, 16, 14, 21, 13, 9, 27, 18, 25, 20, 6

- Find the 35th percentile
  1. Sort the data:
     $6, 9, 13, 13, 13, 13, 14, 14, 16, 18, 18, 20, 21, 25, 27$
  2. Compute the index value $i = \frac{35 \times 15}{100} = 5.25 \Rightarrow$ the 35th percentile is 13

- Find the 65th percentile
  1. Sort the data:
     $6, 9, 13, 13, 13, 13, 14, 14, 16, 18, 18, 20, 21, 25, 27$
  2. Compute the index value $i = \frac{65 \times 15}{100} = 9.75 \Rightarrow$ the 65th percentile is 18

Z-scores & Empirical Rule

- Z-score:

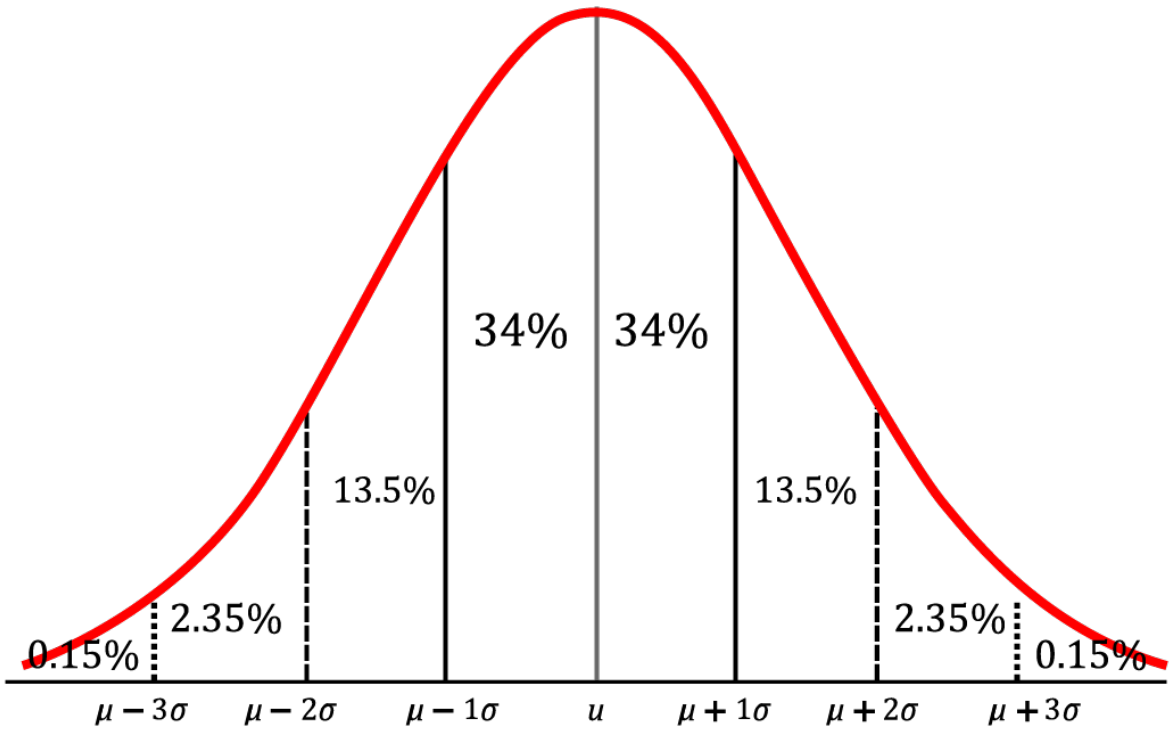
$$z = \frac{x - \bar{x}}{s},$$

when x is the value of an individual observation, $\bar{x}$ sample mean, and s sample standard deviation

- Measuring “how far" (in terms of standard deviations) an observation is from its mean (e.g., 3 standard deviations above the mean value)

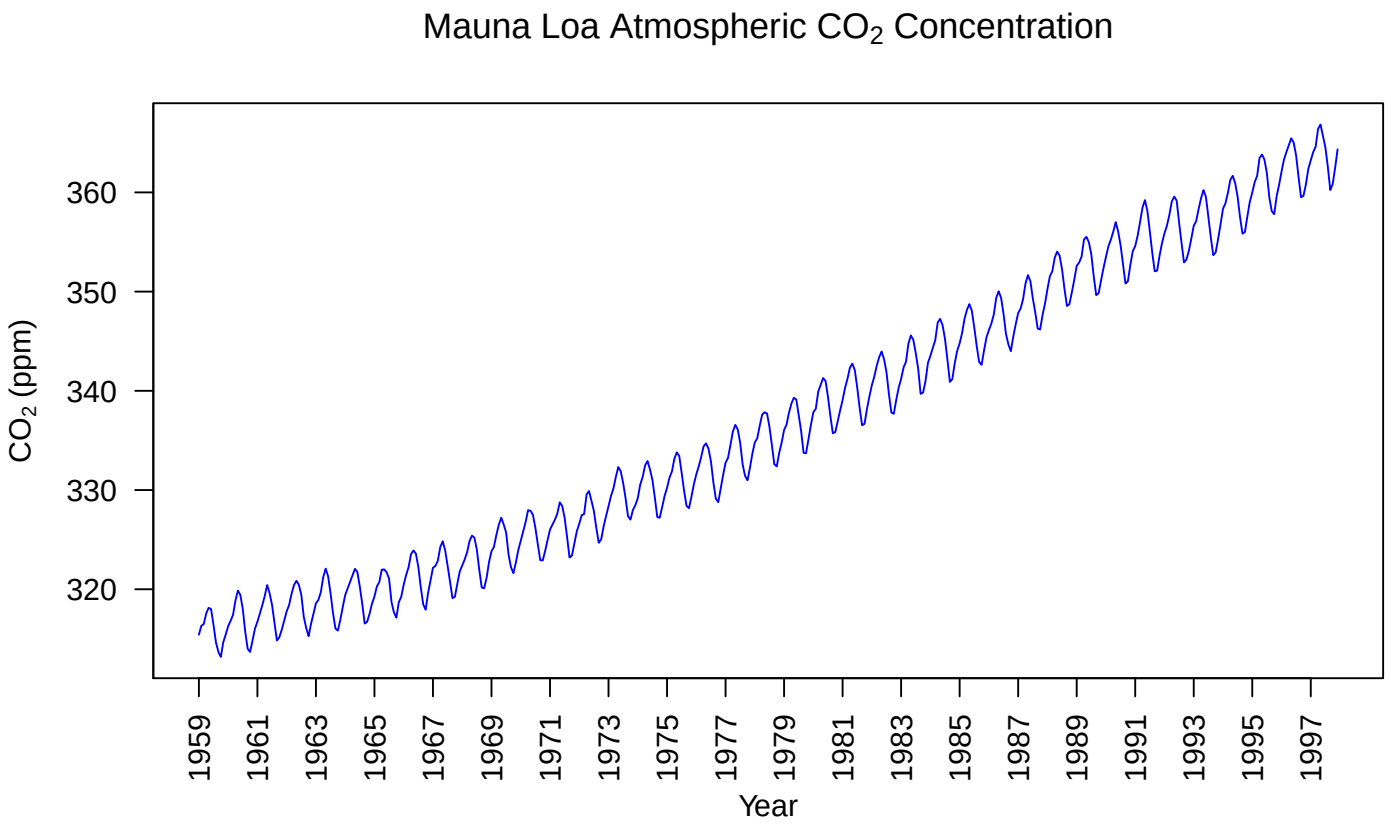
- Empirical Rule: If a data set **can be well approximated by a normal curve (bell-shaped with light tails)**, then approximately 68%, 95%, and 99.7% of the observations are within 1, 2, and 3 standard deviations of the mean

Norm (Bell-Shaped) Curve

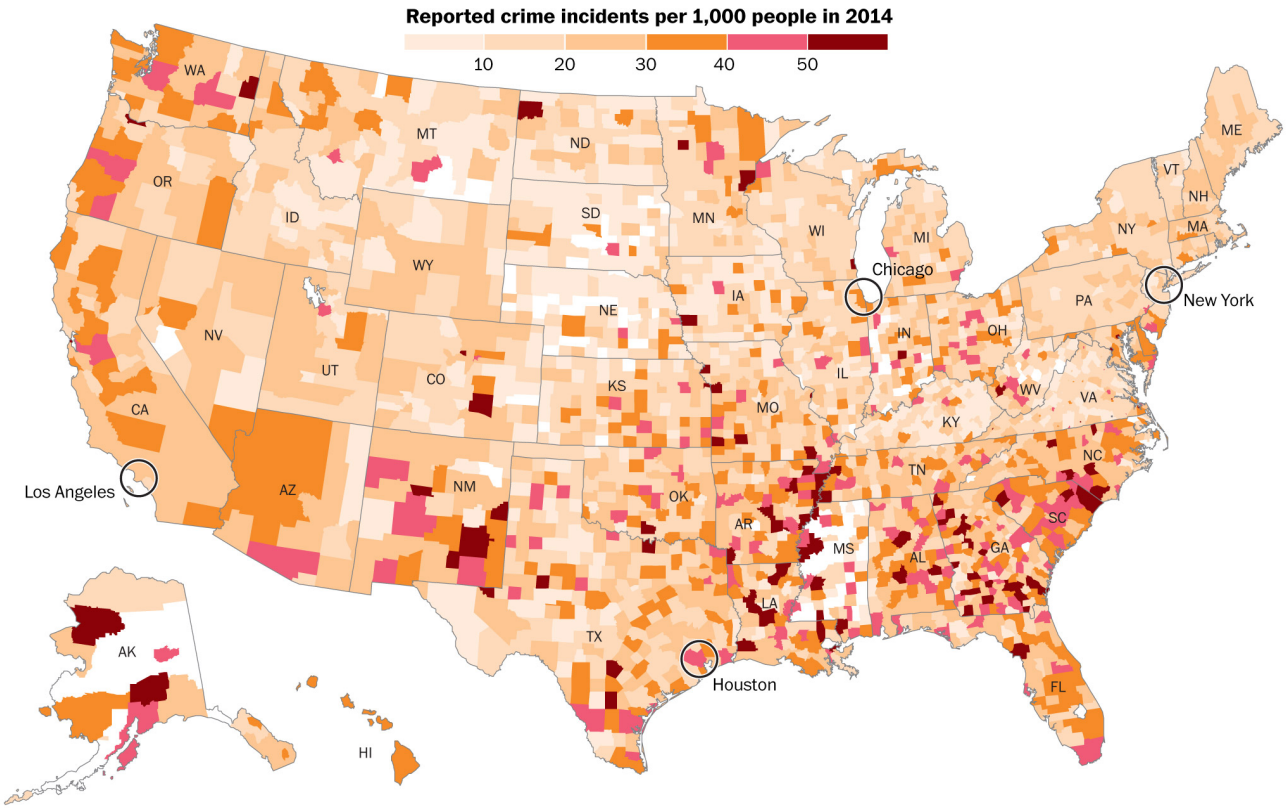


CLEMSON UNIVERSITY

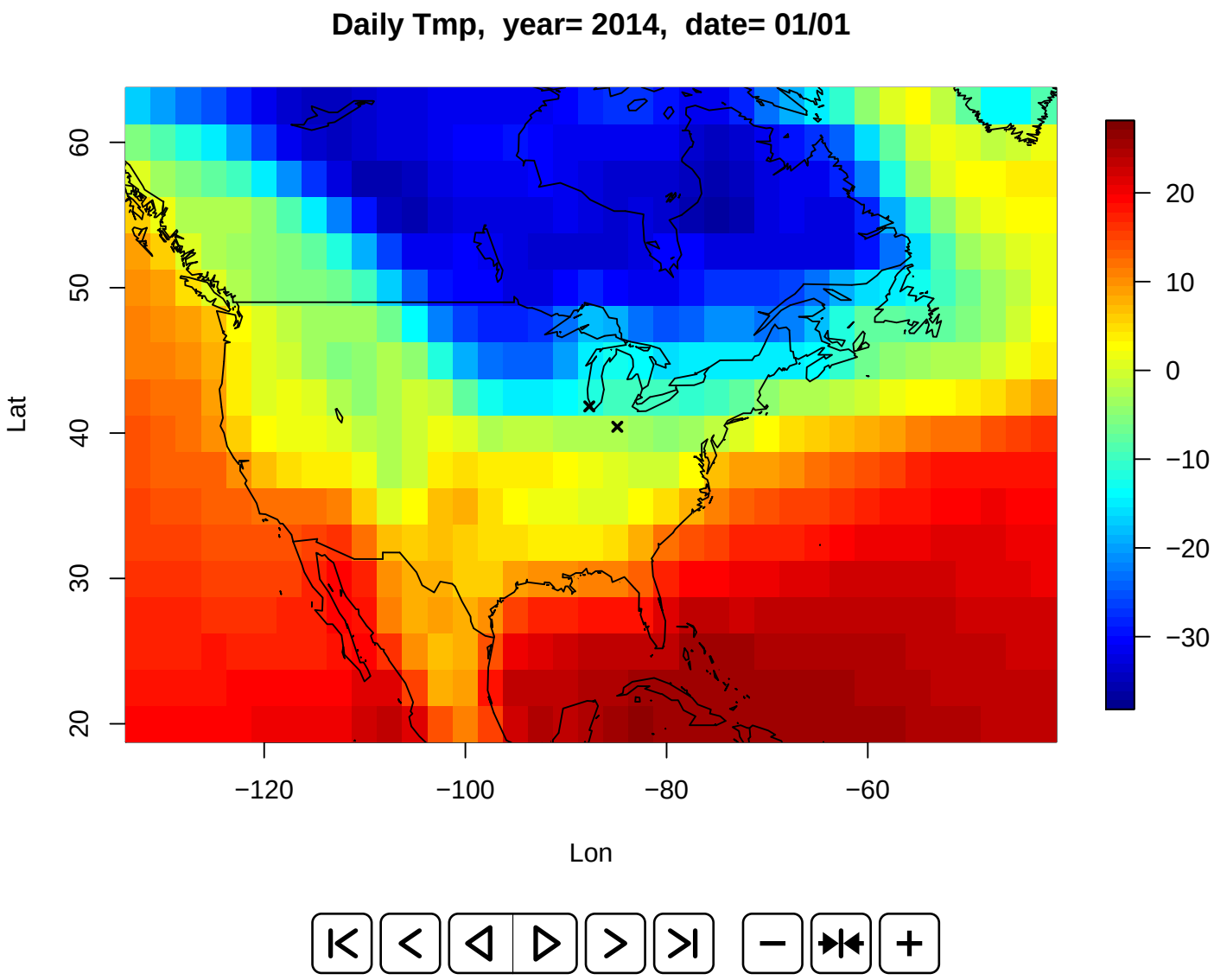
Visualizing Time Series Data



Visualizing Cross-Sectional Data



Visualizing Spatio-Temporal Data



CLEMSON UNIVERSITY

Summary

In this lecture, we learned

- Percentiles and Quartiles
- **How to construct a Boxplot**
- Z-scores & Empirical Rule
- How to visualize time series, cross-sectional, spatio-temporal data sets

We will talk about Probability in next three weeks