



UNIVERSITY OF  
MICHIGAN

# Quantitative Data: Histograms

---

Reed Coots

Statistics with Python Course Developer



© 2018 The Regents of the University of Michigan  
Except where otherwise noted, this work is licensed under  
<http://creativecommons.org/licenses/by-nc/3.0/>

# What are Quantitative Variables?

Variables that have a numerical value (quantity) that we can perform mathematical operations on

**Examples:** Height, weight, income, test scores, shoe size, number of “heads” after 10 coin flips

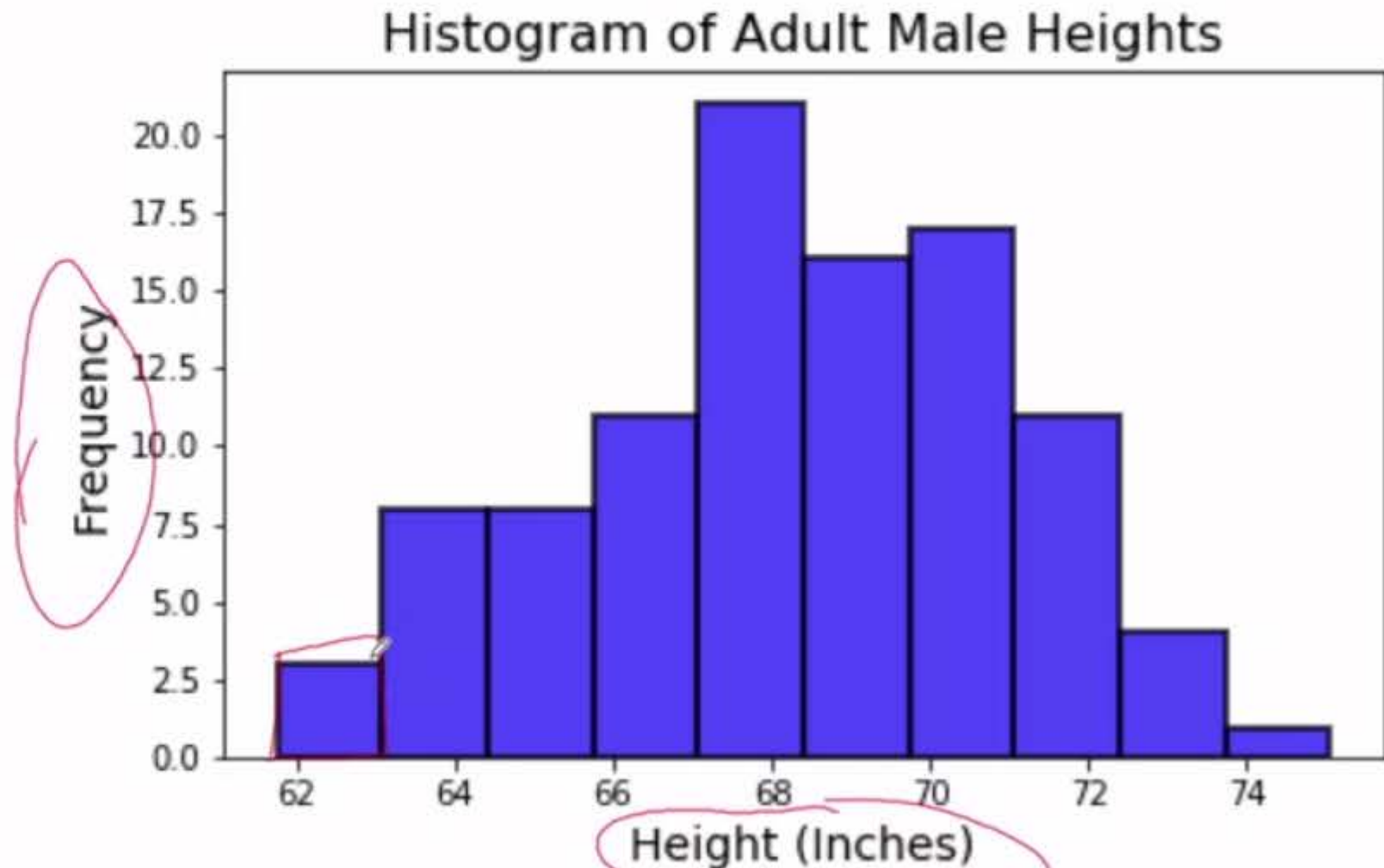


five-foot-seven and a half,  
any value between these numbers.

# Why Use Histograms?

## Adult Male Heights

66.3
75.1
67.9
67.6
70.0
69.9
64.8
...



And then we have all this rectangles  
which are called bins, and



## 4 Main Aspects

- Shape -** Overall appearance of histogram. Can be symmetric, bell-shaped, left skewed, right skewed, etc
- Center -** Mean or Median
- Spread -** How far our data spreads. Range, Interquartile Range (IQR), standard deviation, variance.
- Outliers -** Data points that fall far from the bulk of the data  
Outliers are the final thing  
that we want to talk about,

# Adult Male Heights

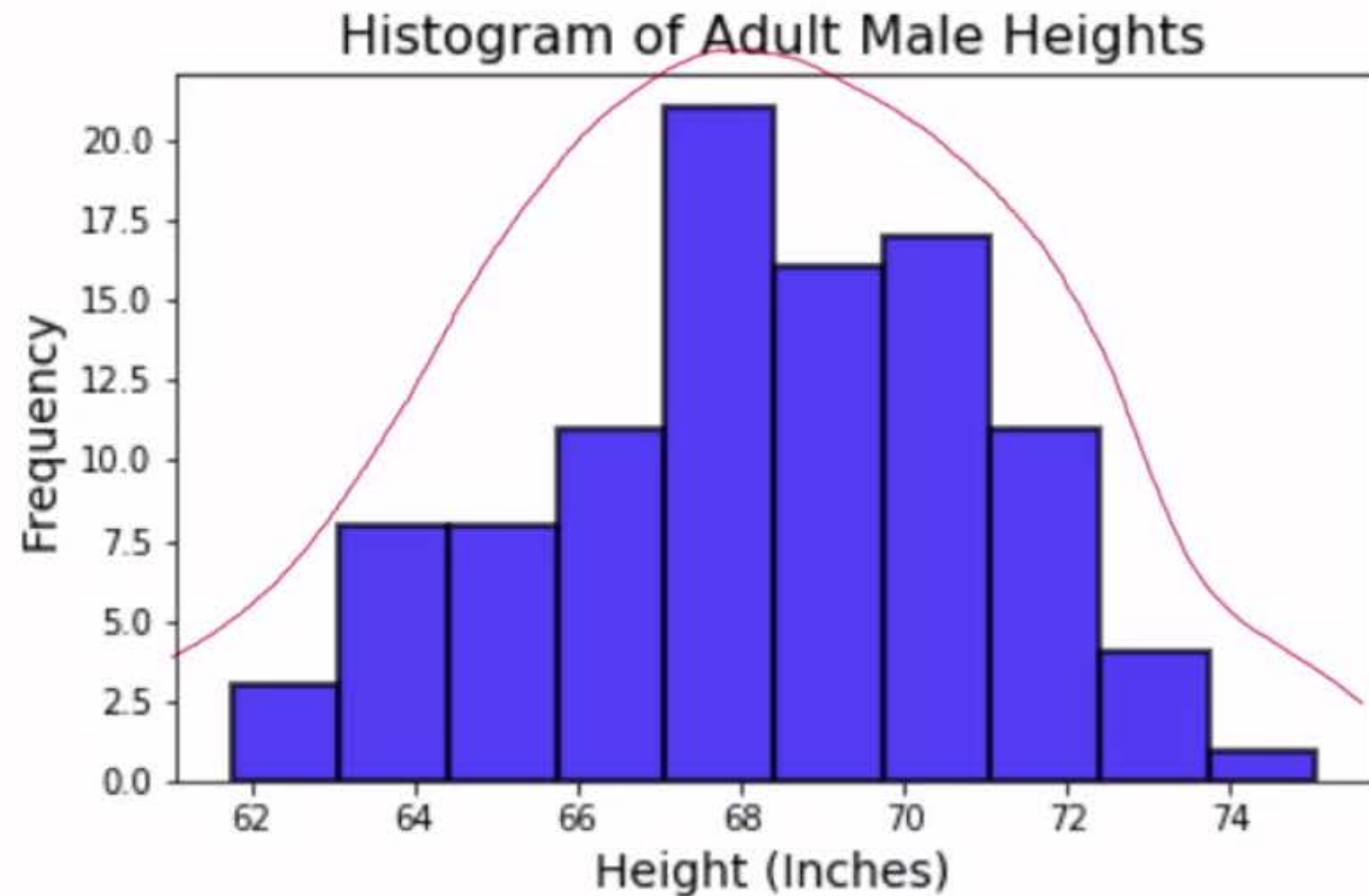
**Shape**

*Bell-shaped*

**Center**

**Spread**

**Outliers**



We can also say that this is unimodal,  
meaning it only has one mode,

# Adult Male Heights

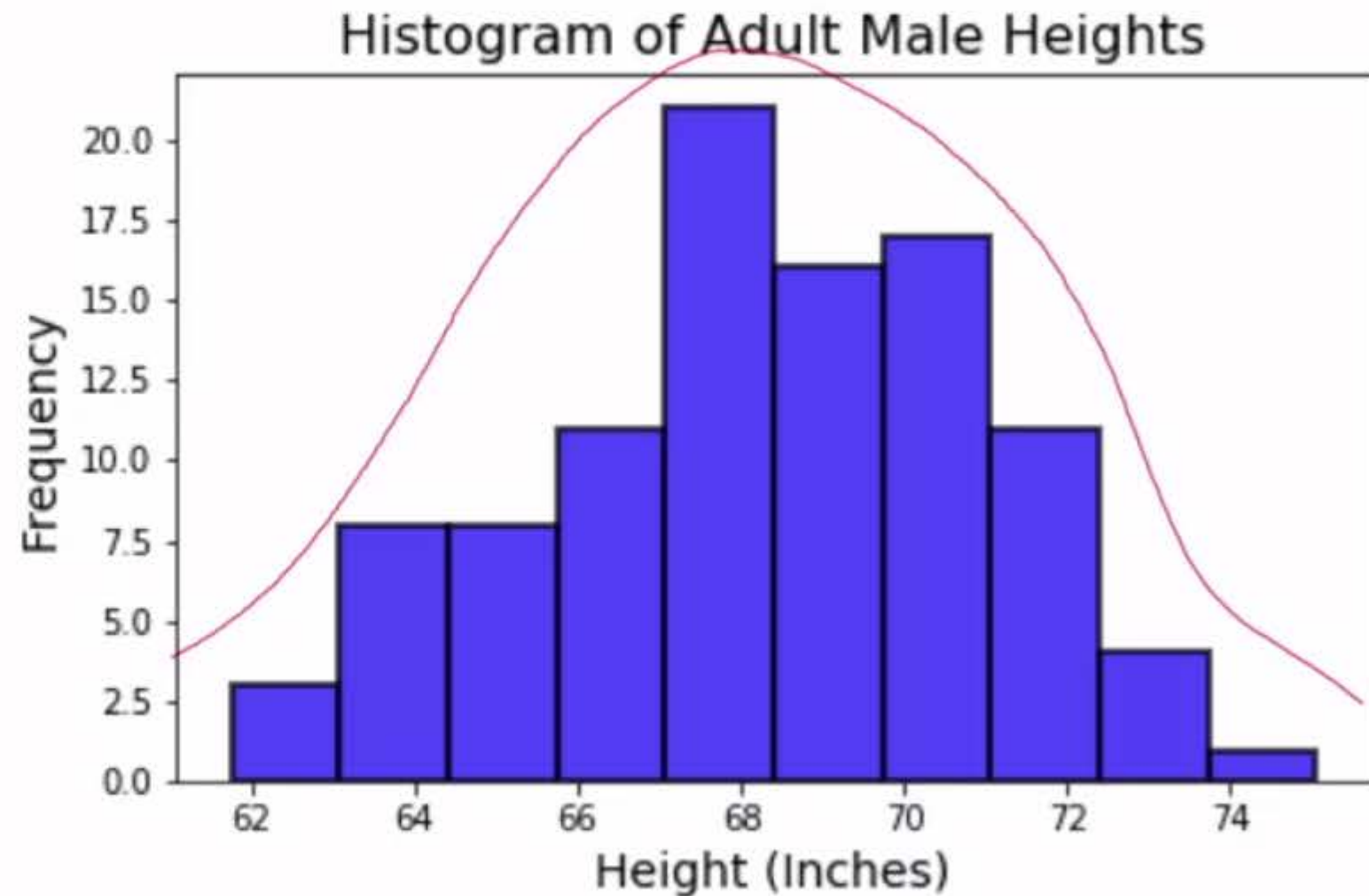
**Shape**

*Bell-shaped  
Unimodal*

**Center**

**Spread**

**Outliers**



and that's because we only  
see one peak in our shape.



# Adult Male Heights

**Shape**

*Bell-shaped  
unimodal*

**Center**

*Median = 68  
Mean = 68*

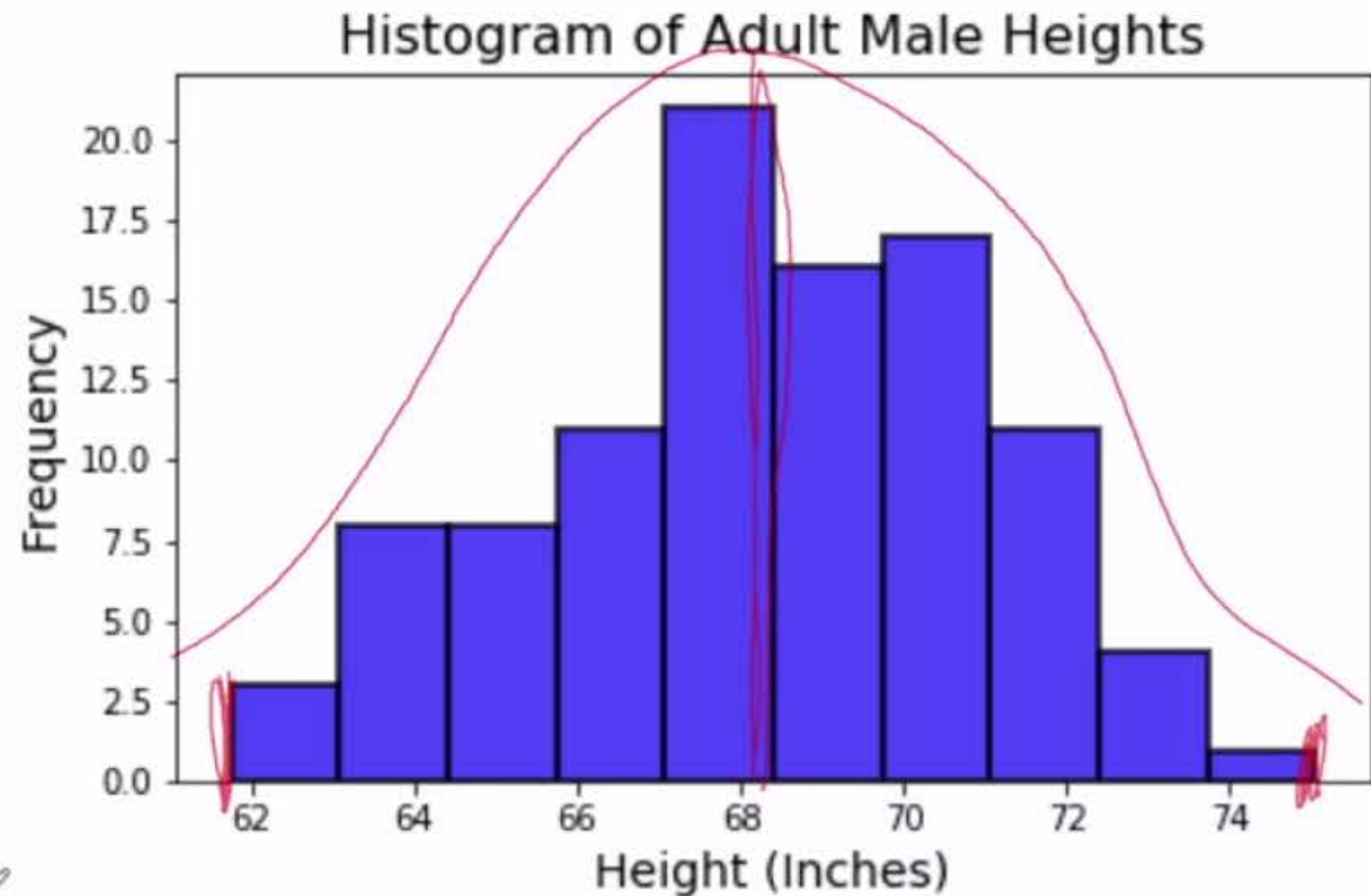
**Spread**

*Range = max - min  
75 - 62  
13*

**Outliers**

*No Apparent outliers*

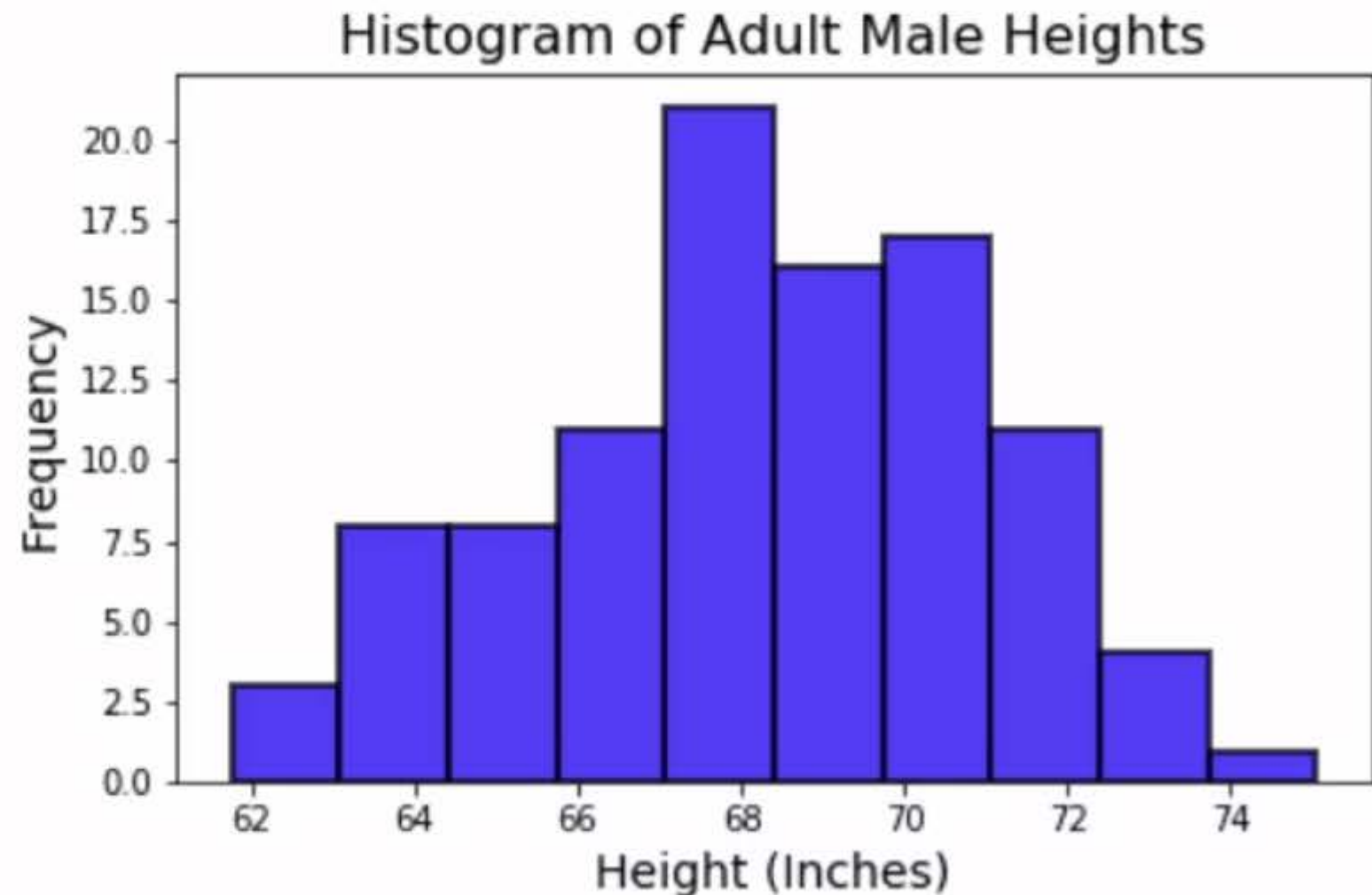
**concise manner.**



# Adult Male Heights

## Putting it all together:

The distribution of adult male heights is roughly bell shaped with a center of about 68 inches, a range of 13 inches (62 to 75), and no apparent outliers.



And anything that they really  
want to know about it.



# Salaries in San Francisco (2011-2014)

**Shape** Right Skewed  
Bimodal

**Center** median = \$80,000  
mean = \$85,000

**Spread** Range = \$600,000

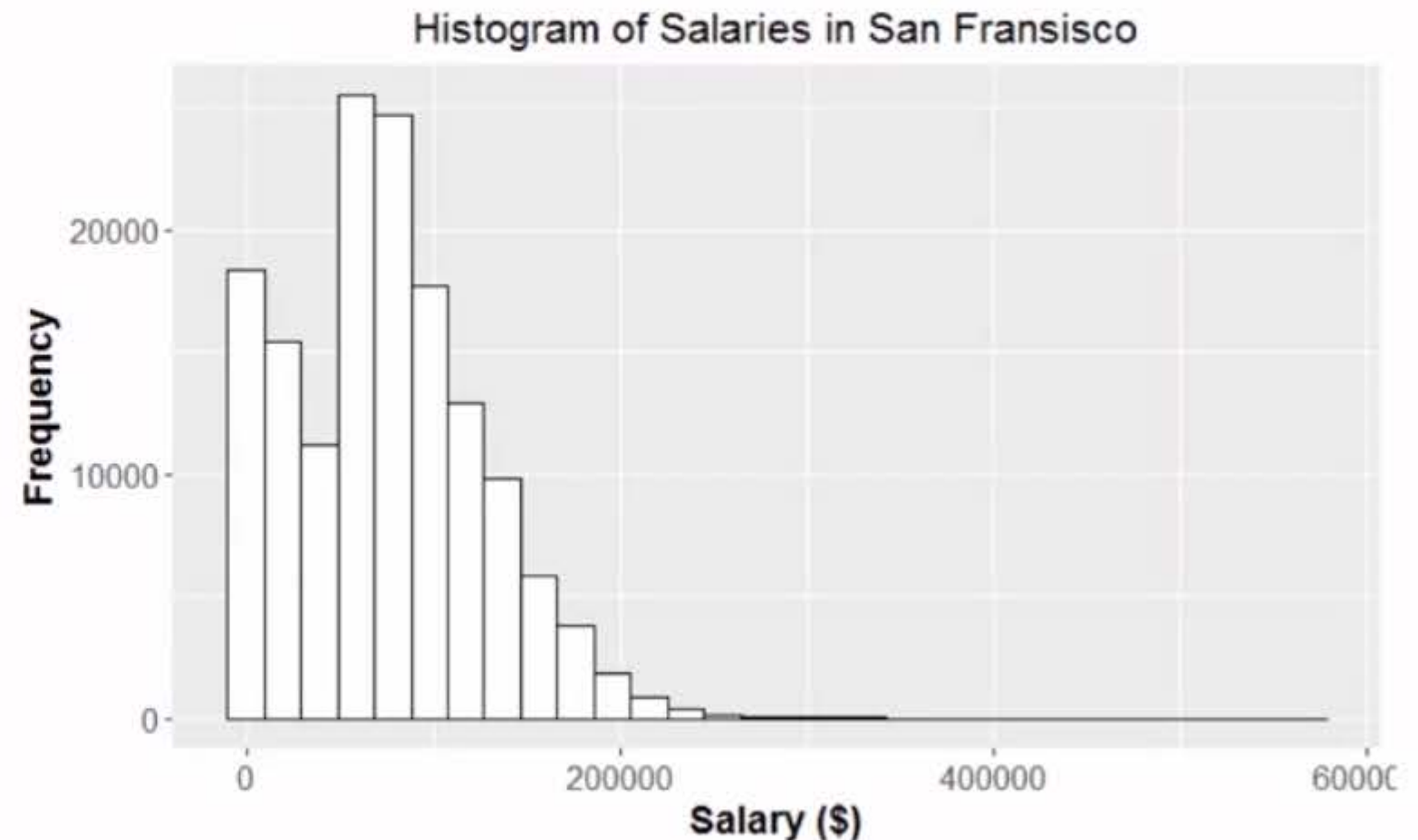
**Outliers** High End



Source: <https://www.kaggle.com/kaggle/sf-salaries/data>

# Salaries in San Francisco (2011-2014)

The distribution of salaries in San Francisco is bimodal and skewed to the right, centered at about \$80,000 with most of the data between \$40,000 and \$120,000, a range of roughly \$600,000, and outliers are present on the higher end.

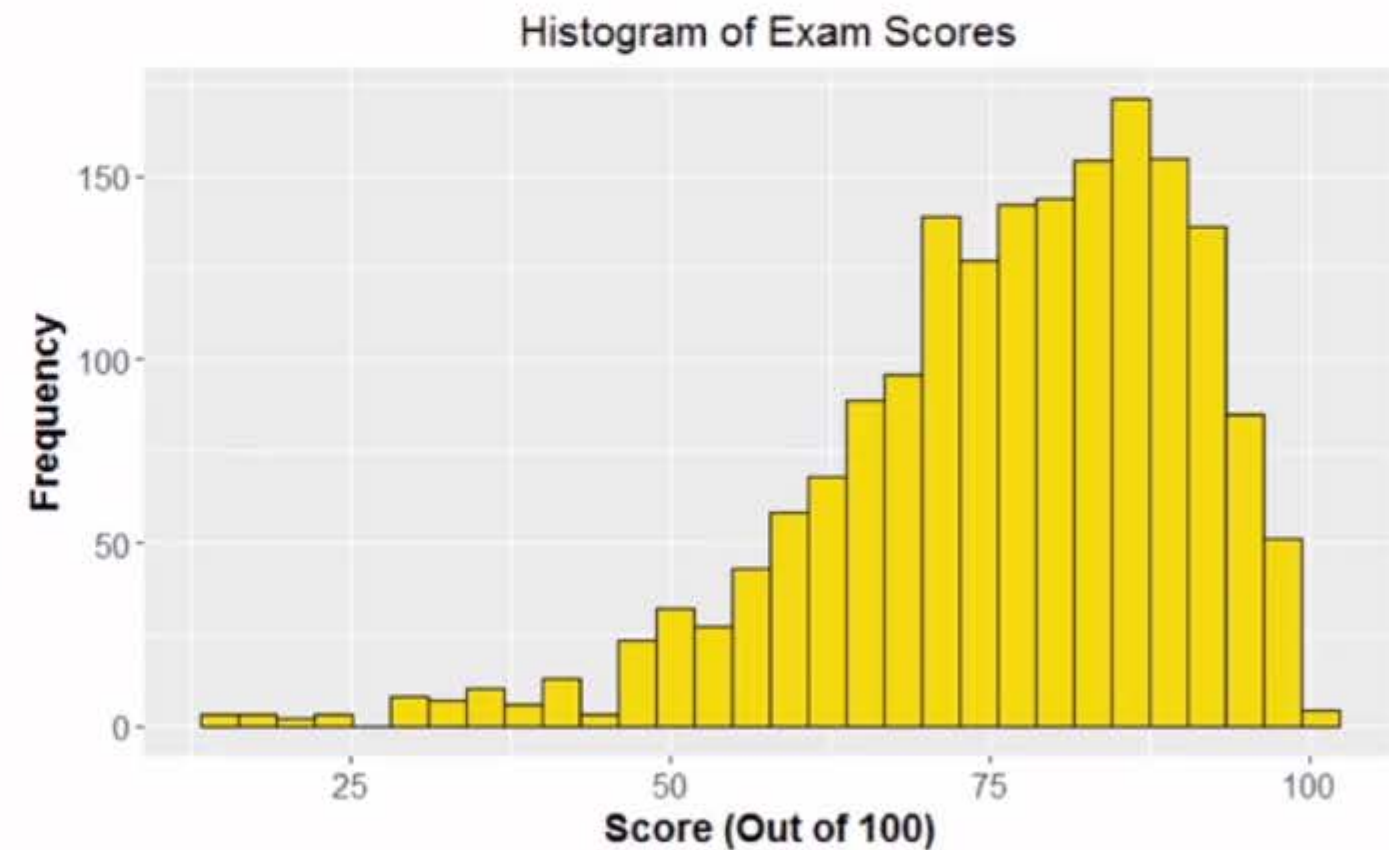


Source: <https://www.kaggle.com/kaggle/sf-salaries/data>



# Exam Scores

The distribution of exam scores is skewed left, centered at about 80 points with most scores being between 65 and 90 points, a range of roughly 85, and some outliers are present below 50 points.



We have a range of roughly 85 and some of the outliers are present below 50 points.



# Summary

- Histograms allow us to display data graphically
- 4 main aspects we use to describe the data
  - **Shape**
  - **Center**
  - **Spread**
  - **Outliers**
- Your one sentence summary should allow for any person to read it and have a general understand of what your data looks like

And with these four things,  
we construct a one sentence summary.