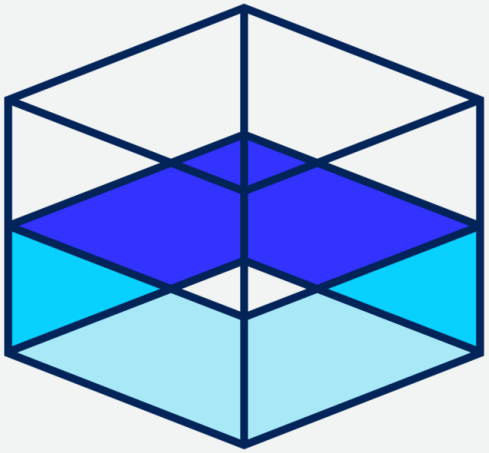


TECH
TALENT
SOUTH

Intro to Data Science

TECH
TALENT
SOUTH

Updated: 10/09/20

Table of Contents

Text

Intro to Data Science

Class Introductions

Introduce yourselves! Discuss:

Location

Current Position

Work History

Education and Training

Expectations for the course

TECH TALENT SOUTH

Intro to Data Science

What is this Course?

During this course you will learn:

- To understand data types, formats, and sources
- Applly statistics to describe data and make projections
- use Excel to perform data analysis
- Design, implement, and query Relational Databases
- Use a variety of charts/graphs to analze and communicate data
- Write Python programs to access, process, analyze, and visualize data
- Work with Jupyter Notebooks
- Gain awareness in topics such as R programming, data mining, Power BI, Tableaeu, NoSQL, and Machine Learning

TECH TALENT SOUTH

Intro to Data Science

TECH TALENT SOUTH

Intro to Data Science

What is this Course?

During this course you will learn:

- To understand data types, formats, and sources
- Applly statistics to describe data and make projections
- use Excel to perform data analysis
- Design, implement, and query Relational Databases
- Use a variety of charts/graphs to analze and communicate data
- Write Python programs to access, process, analyze, and visualize data
- Work with Jupyter Notebooks
- Gain awareness in topics such as R programming, data mining, Power BI, Tableaeu, NoSQL, and Machine Learning

TECH TALENT SOUTH

Intro to Data Science

Roles on Data Science Team

Data Scientist

Data Engineer

Machine Learning Engineer

Data Architect

Business Analyst/Domain Experts

Software Engineer

TECH
TALENT
SOUTH

Intro to Data Science

Roles on Data Science Team

Data Scientist

Data Engineer

Machine Learning Engineer

Data Architect

Business Analyst/Domain Experts

Software Engineer

TECH
TALENT
SOUTH

Intro to Data Science

Data Science Tools

Microsoft Excel

- Tabular data
- Worksheets
- Built-in charts
- Macros

SQL

- Used to work in relational databases

TECH TALENT SOUTH

Intro to Data Science

Data Science Tools

Python

- General-purpose programming language
- Many useful libraries, including those for data science

R

- Language specifically purposed for data analytics/statistics

MS PowerPI and Tableau

- Popular data visualization tools

TECH
TALENT
SOUTH

Intro to Data Science

Where Does Data Come From?

- Functional Area Support Systems
- Corporate Databases
- Government Websites
- Commericial Providers
- Academic/Research Institutions
- Myriad Electronic Devices (IoT)
- DIY

Intro to Data Science

Data Formats

- Flat Files - text-based databases (e.g. csv, tab delimited, JSON, etc.)
- XML Files
- Relational Data
- Unstructured Data

TECH
TALENT
SOUTH

Intro to Data Science

Chart Types

Most of the time, we want to make data a bit easier to read! Let's go over some basic chart types and their uses.

TECH TALENT SOUTH

Intro to Data Science

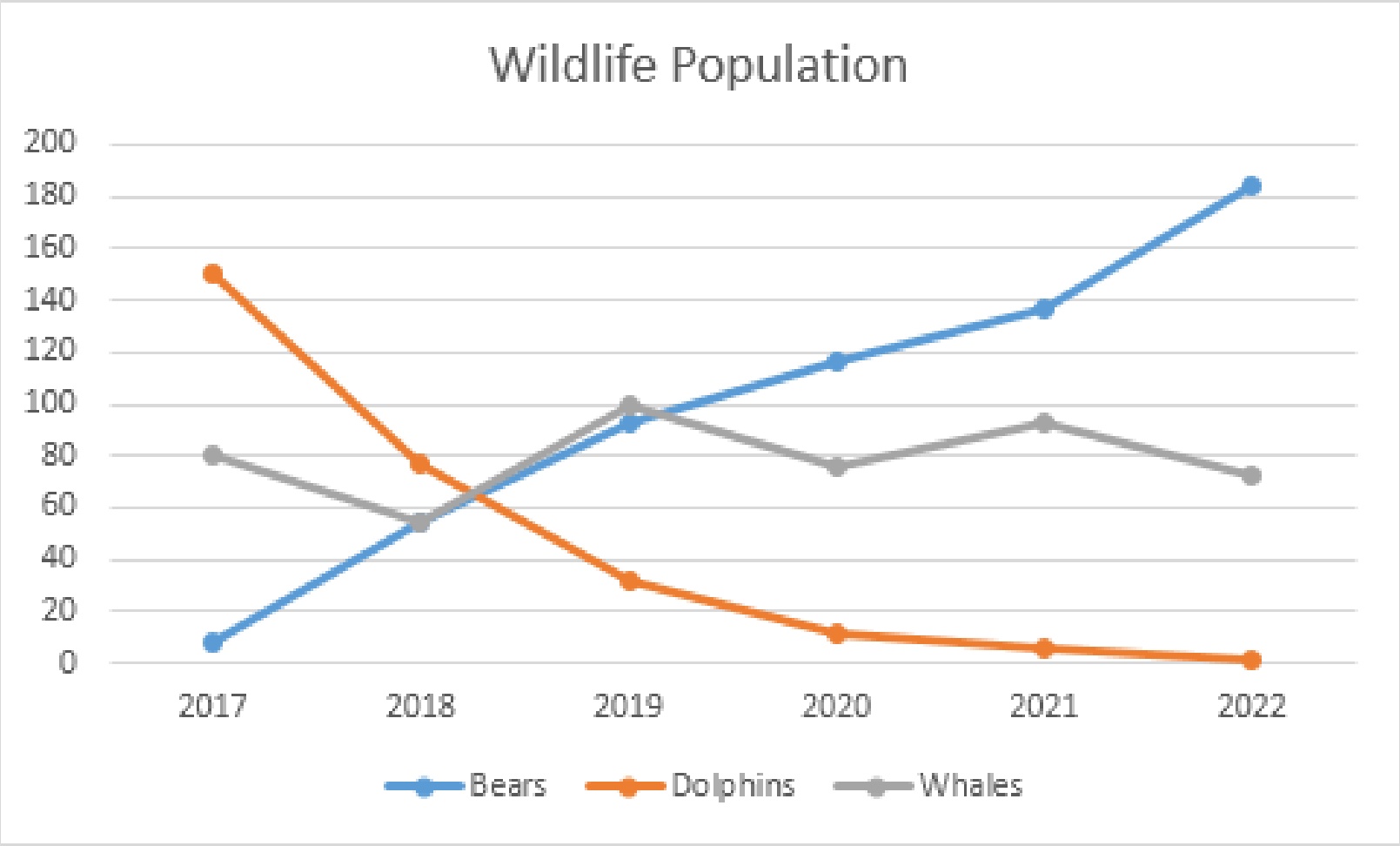
Line Chart

Line charts show continuous data over time on an evenly scaled axis. They are ideal for showing trends in data at equal intervals, such as months, quarters,or years.

In a line chart, category data is distributed evenly along the horizontal axis and value data is distributed evenly along the vertical axis.

TECH TALENT SOUTH

Intro to Data Science

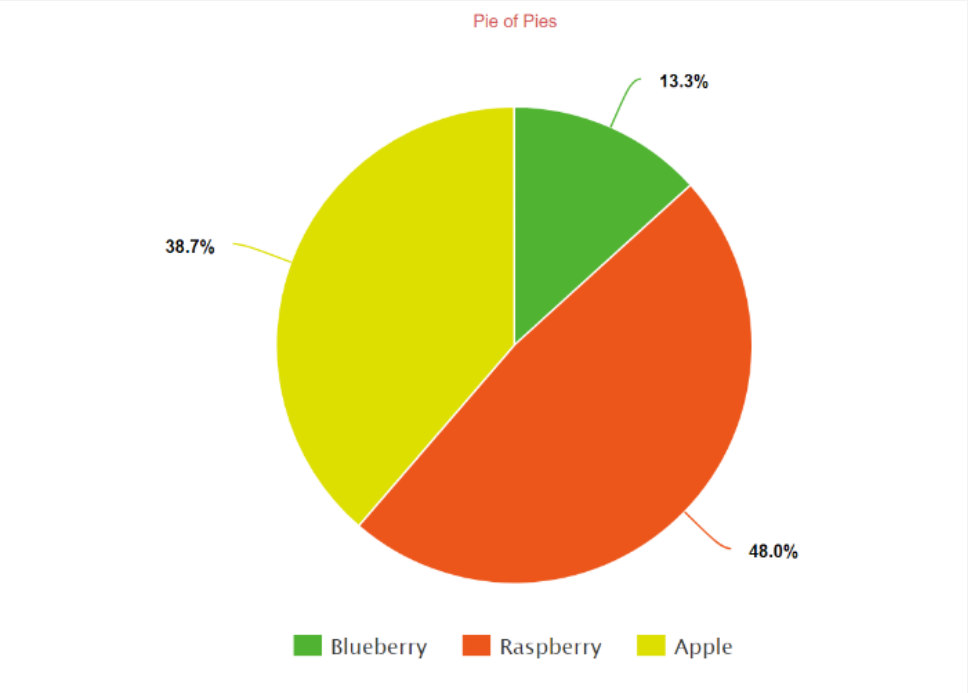


TECH TALENT SOUTH

Intro to Data Science

Pie Chart

Pie charts show the size of items in one data series proportional to the sum of the items.

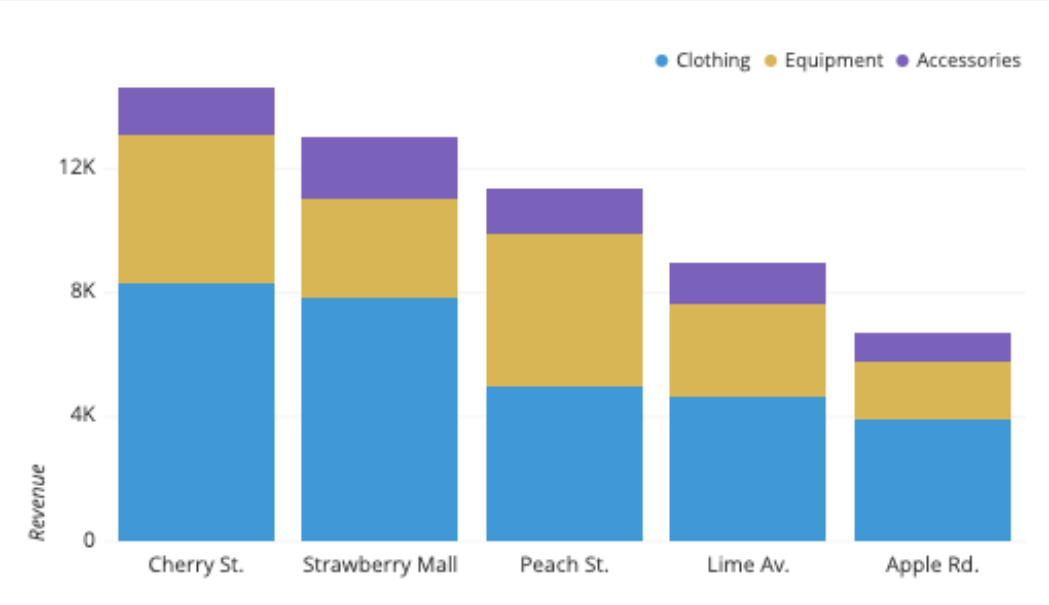


TECH TALENT SOUTH

Intro to Data Science

Bar Chart

Bar charts illustrate comparisons among individual items. The bar chart has a few sub-types, including clustered bar and stacked bar.



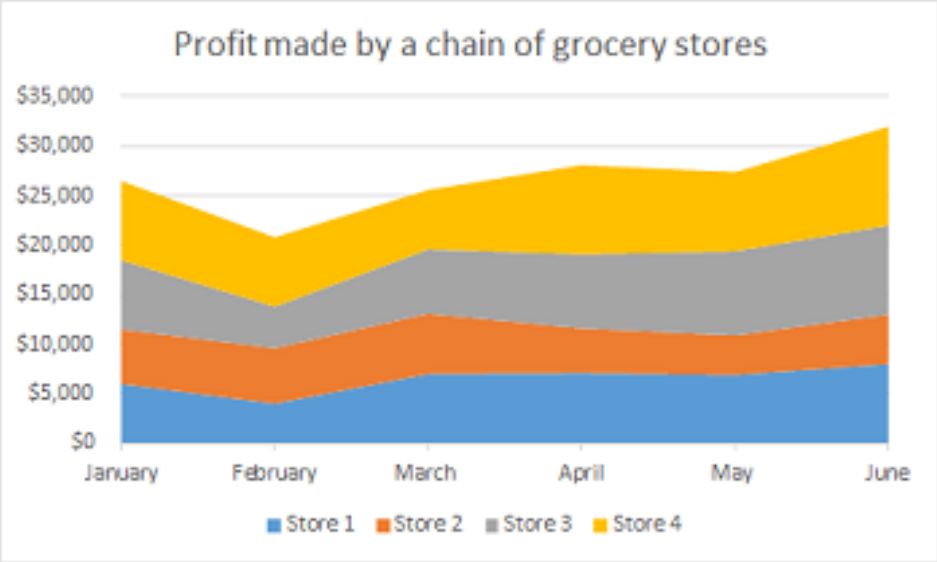
Stacked Bar Chart

TECH TALENT SOUTH

Intro to Data Science

Area Chart

An area chart is a line chart with the areas below the lines filled in. They can be used to plot change over time and draw attention to the total value across a trend. By showing the sum of the plotted values, an area chart also shows the relationship of parts to a whole.

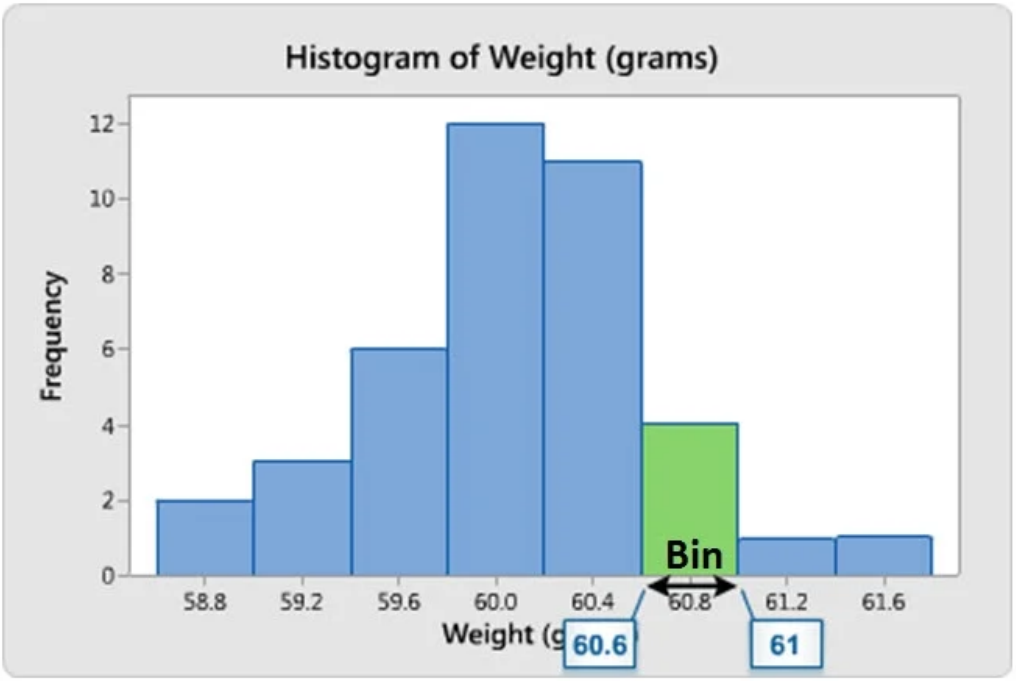


TECH TALENT SOUTH

Intro to Data Science

Histograms

Histograms show distributions of variables. Histograms plot quantitative data with ranges of the data grouped into bins or intervals.

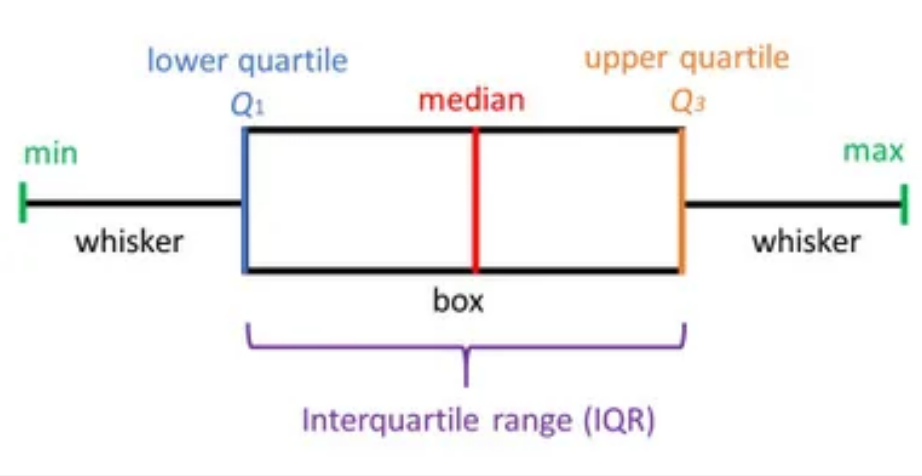


TECH TALENT SOUTH

Intro to Data Science

Box and Whisker

A box and whisker plot displays the five-number summary of a set of data. The five-number summary is the minimum, first quartile, median, third quartile, and maximum.

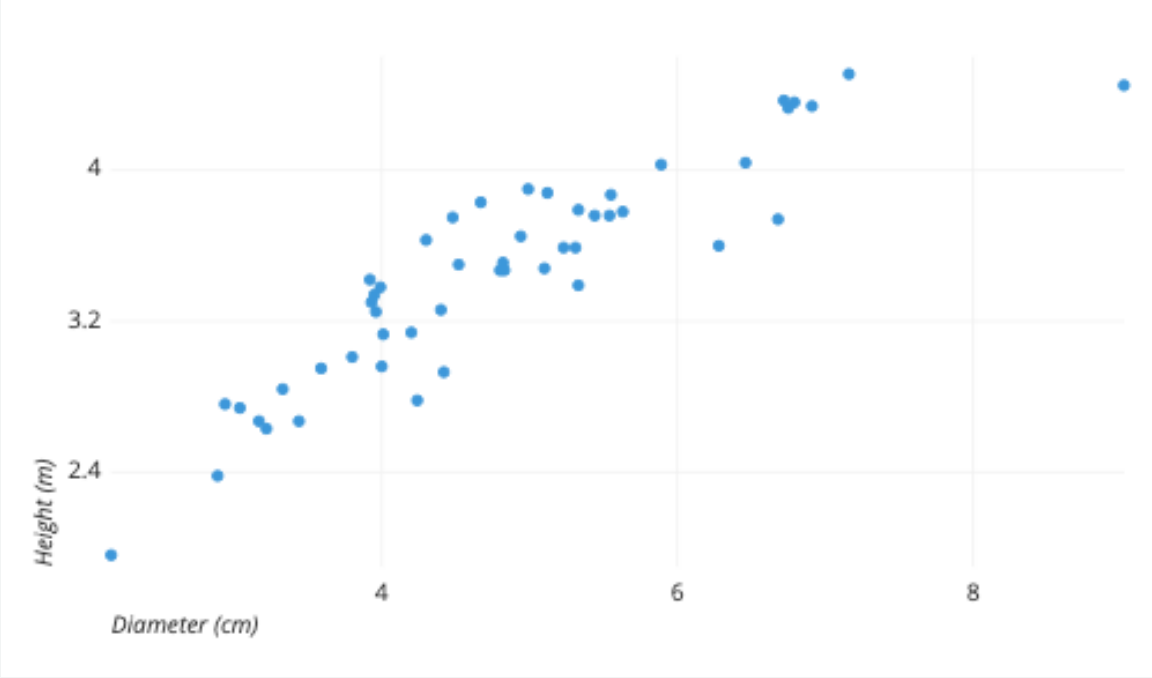


TECH TALENT SOUTH

Intro to Data Science

Scatter Plot

A scatter plot shows scientific XY data. Scatter plots are often used to find out if there is a relationship (correlation) between variable X and Y.



TECH TALENT SOUTH