Mapping Research Requirements to Software Tools:
A Data Science Perspective

Virtual Residency Intermediate/Advanced Workshop 2020
Fitzroy Nembhard, Ph.D.

Harris Institute for Assured Information
http://hiai.fit.edu
About Florida Tech

- Founded in 1958, Florida Tech is a national, private, not-for-profit, technological, coeducational, doctoral-granting, research-intensive university.

- One of America’s most diverse universities, Florida Tech has more than 60,000 alumni from more than 120 countries.

- Three astronaut faculty members: Buzz Aldrin, Sc.D., Samuel Durrance, Ph.D., and Winston Scott, M.S.

- According to the 2020 rankings from U.S. News & World Report, Florida Tech is No. 43 in the U.S. among universities with a full-time undergraduate population of 5,000 students or fewer.

A Data Science Perspective

• **Data science** is a multidisciplinary field focused on finding actionable insights from large sets of raw and structured data.

• **Data Analysis** is a process for obtaining raw data and converting it into information useful for decision-making by users.
Data Science at Florida Tech

Data Analysis Methods Course, Spring 2020

Responses = 16/20 students

```
print("Hello World!")

Learn Python
Get the Data
Preprocess it
Model it, Analyze,
Visualize & Present
```
Data Analysis: Tools for Conducting Surveys

- SurveyMonkey
- FormSite
- SurveyShare
- Google Docs
- eSurveysPro
- ThesisTools Online Surveys
- Free Online Surveys
- Infopoll
- FormSite
- LimeSurvey (formerly PHPSurveyor)
- Kwik Surveys
- QuestionPro
- MailChimp
- Recorded Webinars
- Polldaddy
- Survey Gizmo
- eSurv
Data Analysis: Python Tools

- Beautiful Soup
- lxml
- tweepy
- Pandas
- Panide
- NLTK
- SciKit
- Matplotlib

- Numerical computing in Python
- Numpy
- Scipy
- Sympy
- Staionmodels
- Sciklearn
- Tensorflow

- Get Data
- Preprocessing
- Modeling & Analysis
- Evaluate, Visualize & Present

- Data Analysis Tools
Data Analysis: R Programming Tools

Data Analysis Tools

Mapping Research Requirements to Software Tools
Other Data Analysis Tools

- KNIME
  - Konstanz Information Miner, open-source data analytics, reporting and integration platform: ML, odbc, etc

- Tableau Public
  - Connect to a spreadsheet or file and create interactive data visualizations for the web

- SAS
  - Developed by SAS Institute for data management, advanced analytics, multivariate analysis, business intelligence, criminal investigation, and predictive analytics.

- Spark Streaming
  - Spark SQL
  - Machine Learning
    - Mlib
    - GraphX

- Apache Spark

- Splunk

- Excel

- IBM SPSS Modeler

- "Other"

Mapping Research Requirements to Software Tools

© Fitzroy Nembhard
Big Data Analytics Tools

- Hadoop: Data Storage and Processing Platform
- Kafka: Messaging System
- Hive: Data Warehousing
- Spark: Realtime Processing
- Talend: Software Integration Platform
- NoSQL Database
- Amazon DynamoDB
- Apache HBase
- Apache Spark
- Splunk: Log Analysis Platform

© Fitzroy Nembhard

Mapping Research Requirements to Software Tools

Image Recreated from: https://www.houseofbots.com/
Thank You!  http://www.fit.edu