Scope:

A foundation of knowledge in the data sciences can prove advantageous for almost all scientists. Often, however, understanding the fundamentals of any new skill lacking proper instruction or recourses can be difficult. With ample guidance, all can develop a base knowledge of Data Science that can grow and provide new skills for improved research experiences.

The sets of instruction, examples, and tutorials in this walk-through are specifically designed for Merck Technology and can be distributed to Merck Scientists. The walk-through will be tailored to the needs of the pharmaceutical industry and will relate to work being done at Merck.

Objectives:

1. Data Science Overview
   1. What is data Science?
   2. How is data science used today?
   3. What are the results of data science? – ML/NLP/Deep Learning/projects
2. Getting Acquainted with Merck Technology (Cluster)
   1. Getting Access
   2. Setting up the Client
   3. Experimenting with different resources on the cluster
   4. Introduction to filesystems
3. Navigating through File systems - terminal
   1. Directories
   2. Files
      1. Text Editors – nano/vim
   3. Important terminal commands
      1. pwd, cd, ls, mkdir, rmdir, touch, open, cp, rm, mv, ctrl-c or q, clear, help, whatis, history
   4. Practice Exercise: make new directory, make new text file, add hello world, remove file
4. Introduction to R – Jupyter Notebook or R Studio?
   1. Basic Information about R
   2. Data Frames, Vectors, Indexing, Plotting, subset
   3. Vectorized operations,
   4. tapply, sapply
   5. discussion of popular packages – ggplot, packages
   6. lists?
   7. Practice Exercise – Some basic data frame questions
5. Introduction into Unix/Bash - terminal
   1. Basic info about Unix/Bash
   2. Pipes, awk, grep
   3. Processing speed
   4. Practice Exercise …
6. Introduction to SQL – terminal or mySQL?
   1. Basic Information about SQL
   2. Data frames, tables, columns, rows, queries
   3. Important SQL commands
      1. SELECT, \*, FROM, WHERE, AS, ORDER BY, JOIN
   4. Practice Exercise …
7. Introduction to Python Jupyter Notebook
   1. Basic Information about Python
   2. Variables, Lists, dictionaries, tuples, sets
   3. Loops, conditionals
   4. Functions
   5. Pandas, Numpy, matplotlib
      1. Data frame operations
      2. Plotting
   6. Practice exercise
8. Deeper into R: Shiny and Tidyverse
   1. Intro about Tidyverse
   2. Basics of tidyverse
      1. tibbles, dpylr, tidyr, data.table, ggmap
   3. Basics of Shiny
      1. Inputs and outputs, making an app, reactive expressions, publishing an app
   4. Practice Exercise
9. Deeper in Python: SQL and Object-Oriented Programming
   1. Intro to SQLAlchemy
   2. Queries within python
      1. Loops, sql commands, .notation
   3. Intro to Object Oriented Programing
      1. Classes, dunder methods, methods
10. Intro to Statistics
11. Intro to Probability
    1. Probability Basics
    2. Discrete
    3. Continuous
12. Intro to Linear Algebra
13. Statistical Learning (from “*An Introduction to Statistical learning”* book)
    1. Linear regression and classification
    2. Resampling Methods
14. Intro to Machine Learning
    1. Unsupervised learning
    2. Supervised learning

<https://thedatamine.github.io/the-examples-book/r.html#r-scraping>

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Outline for Biometrics Project:

Scope:

The process to approve a new drug for the public is long and costly. With the emergence of wearable fitness technology, biometric data is being collected quite rapidly and with decent accuracy. By automating a process of collecting data from wearable technology, it will be possible to collect data from clinical trial subjects and help with drug analysis. Long term, this may help speed up clinical trial time and cut costs due to ease of data collection.

Objectives:

1. Introduction to Project
   1. Overall Goals of Project
   2. What is expected from reading this text
   3. Introduction to Fitbit technology
2. Merck Technology
   1. Gaining Access
   2. Setting up the Client
   3. Experimenting with different resources on the cluster
   4. Introduction to filesystems
3. Fitbit Technology
   1. <https://dev.fitbit.com/build/reference/web-api/>
   2. Discussion of general API Basics
   3. In Depth look at API process for Fitbit
      * + 1. App Registration
          2. Authentication
          3. Libraries – OAuth 2.0?
   4. Discussion of Data collected
4. Dependencies
   1. Anaconda Environment
   2. Dependencies List
   3. How to set up anaconda environment
   4. Text Editor – VS code/sublime
5. Working in Python
   1. Python Introduction
      * + 1. Variables
          2. Lists
          3. Dictionaries
          4. Functions
          5. Loops/conditionals
          6. Data Frames
          7. Packages and imports
   2. Walk Through of python script
      * + 1. Necessary imports
          2. Authorization
          3. Functions

getSleep

getActivities

getWeight

getHeart

* + - * 1. Data Frame creation
        2. getBiometricData

dictionaries

* + - * 1. csv file

1. Working in Terminal
   1. Terminal Introduction
      * + 1. File Systems
          2. Directories
          3. Files
          4. Text Editors – nano/vim
          5. Important terminal commands

pwd, cd, ls, mkdir, rmdir, touch, open, cp, rm, mv, ctrl-c or q, clear, help, whatis, history

* + - * 1. Bash Scripts
        2. Unix basics

cat/head

* 1. Using Bash Script to get Biometric Data
     + - 1. Source activate conda env
         2. Call python script
         3. cat files together
  2. Cron

1. Working in R
   1. R introduction
      * + 1. Data Frames, Vectors, Indexing, Plotting, subset
          2. Vectorized operations,
          3. tapply, sapply
          4. discussion of popular packages – ggplot, packages
          5. lists
   2. Analysis of collected biometric data
      * + 1. Simple R exercises?
   3. Shiny
      * + 1. Inputs and outputs
          2. making an app
          3. reactive expressions
          4. publishing an app
2. Working in SQL
   1. SQL Introduction
      * + 1. Databases, tables, columns, rows, queries
          2. Important SQL commands

SELECT, \*, FROM, WHERE, AS, ORDER BY, JOIN

* + - * 1. SQL in terminal and in SQL Workbench
  1. Amendments to Bash Script
     + - 1. Adding to SQL Database

1. Working with Flask
   1. Installations
   2. REST API architecture introduction
      * + 1. Overall Purpose
          2. HTTP requests – GET POST DELETE …
          3. Use in this project
   3. Configuring Database
      * + 1. Yaml file
          2. Host, username, password
   4. Writing search Queries within Python
      * + 1. Cursor
          2. Execute
          3. Commit
          4. Jsonify
2. Working with React Native and Expo
   1. Installations
   2. Expo introduction
      * + 1. …
   3. React Native Introduction
      * + 1. Hooks
          2. Functions
          3. Imports
          4. ….
   4. ….
3. Publishing project
   1. …

```{r}

library(DBI)

db <- dbConnect(RSQLite::SQLite(), dbname = "./imdb.db")

```

```{sql, connection=db}

SELECT \* FROM episodes LIMIT 1;

```