$\text{PS2}_Cocklin$

connorcocklin

January 2019

1 List of a Data Scientist's tools

1. Statistical Programming Languages

   (a) R - biggest advantage is Tidyverse
   (b) Python - biggest advantage is Ubiquity
   (c) Julia - biggest advantage is the speed of computation

2. Handling Large Data Sets

   (a) RDDs-resilient distrinuted datasets
   (b) SQL - most common language to transform data into a more usable form for statistical software to use

3. Visualization

   (a) ggplot2 (R)
   (b) matplotlib (python)
   (c) Plots.jl (Julia)
   (d) Tableau