BIOSTATISTIC PRACTICAL USING R

Adil ZA MBBS, DLSHTM, MSc, MPH, DrPH
DEPARTMENT OF COMMUNITY MEDICINE
KULLIYYAH OF MEDICINE
IIUM

COMMUNITY MEDICINE

STATISTICIAN DO IT WITH **CONFIDENCE**, FREQUENCY AND VARIATION

OBJECTIVE

- Biostatistic practical sessions using R Studio software
- Application for community survey
- Learning mode
  - Series of recorded lecturer (asynchronous mode)
  - Practical exercise (synchronous session)
  - Continuous assessment (TBA, practical assessment)

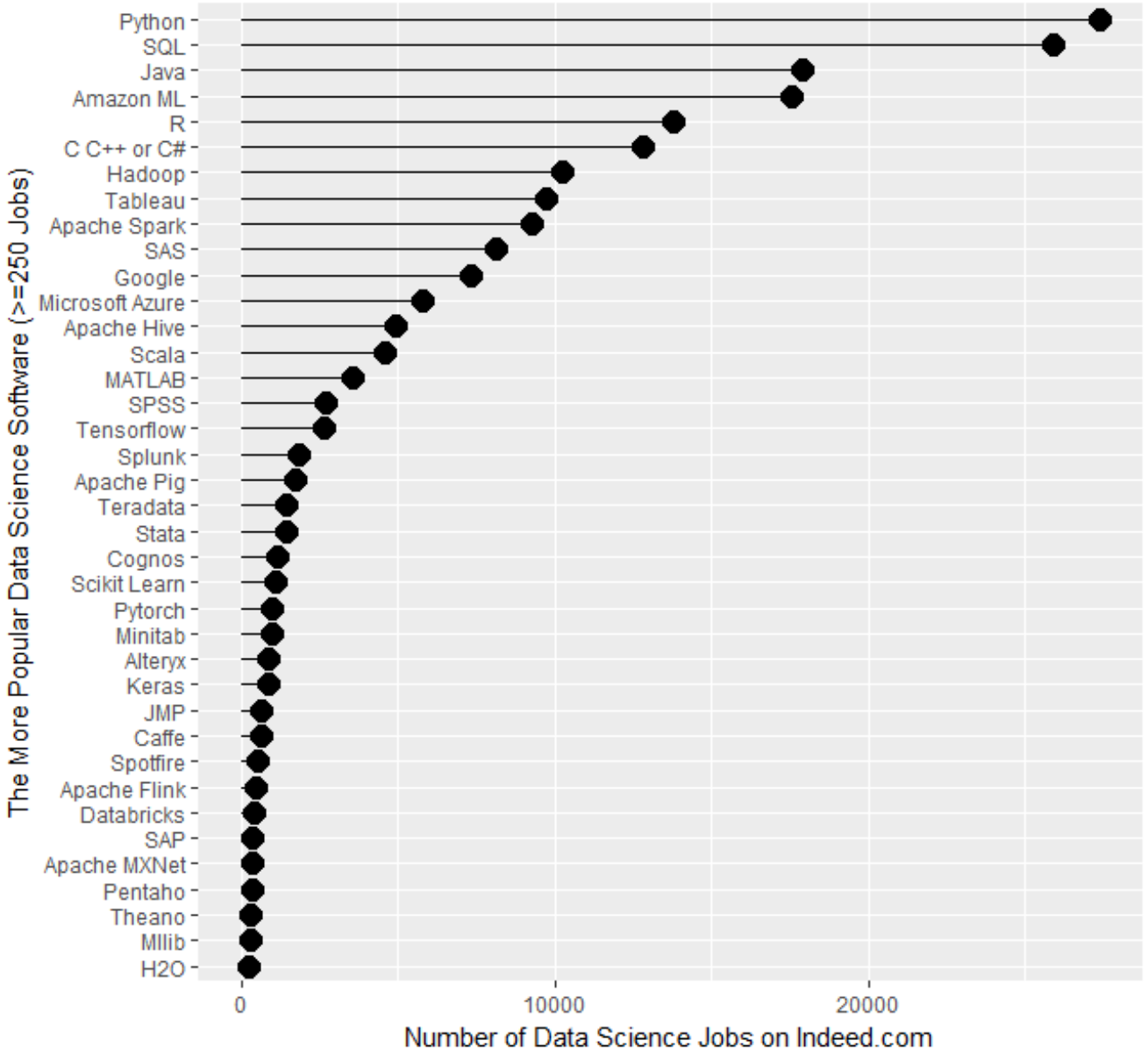
LECTURE TOPICS

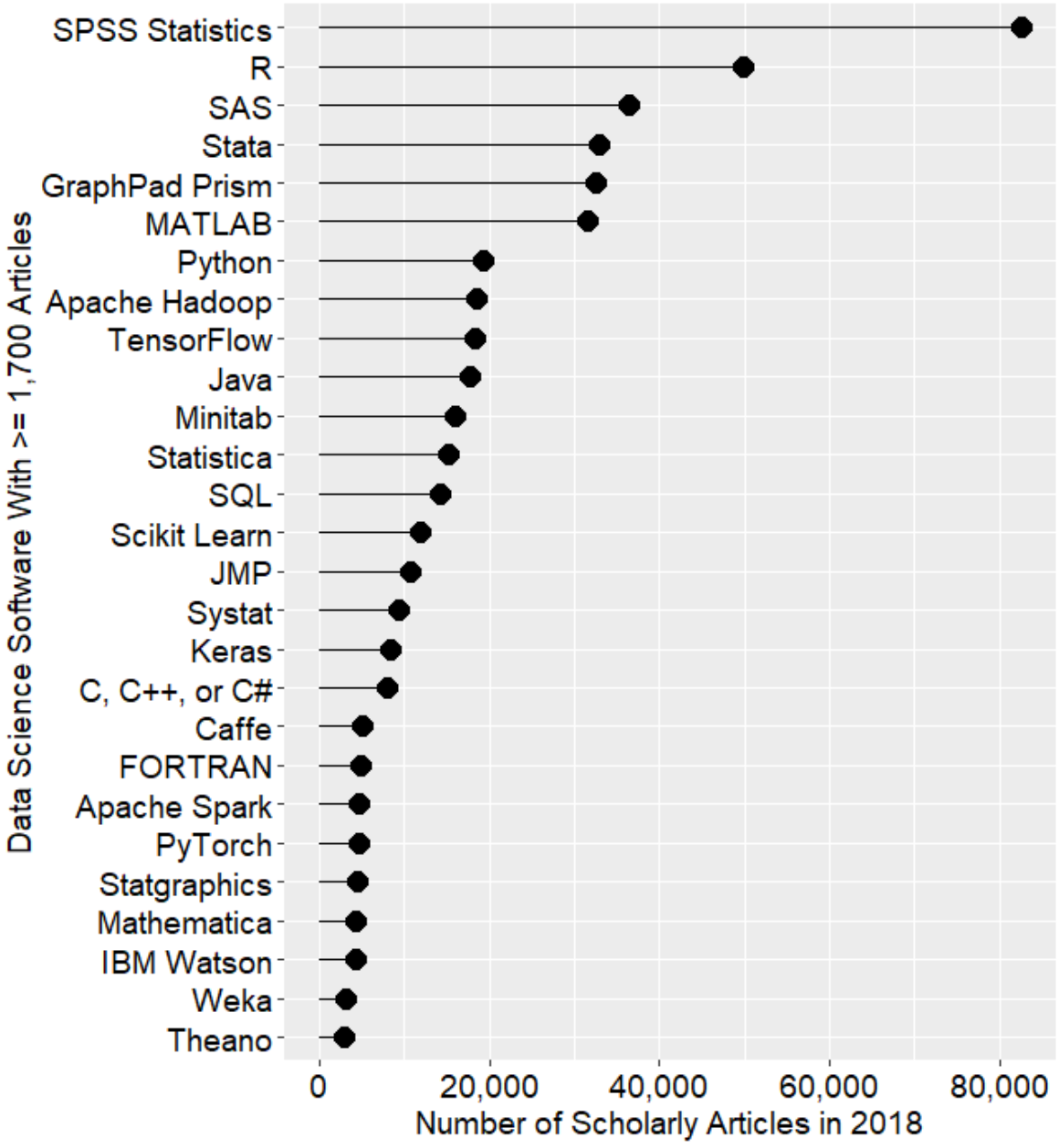
1. Principles of biostatistics (Prof Jamal)
2. Introduction to R
   - Introduction to R (Lecture 1)
   - Data management using R (Lecture 2)
3. Descriptive statistics (Lecture 3)
4. Univariate analysis 1
   - Chi square test (Lecture 4)
   - Correlation test (Lecture 5)
   - Independent t-test (Lecture 6)
5. Univariate analysis 2
   - ANOVA (Lecture 7)
   - Non-parametric test (Lecture 8)
   - Linear and logistic regression (Lecture 9)

LET'S STARt

- Download R software – https://posit.co/download/rstudio-desktop/
- Download dataset https://github.com/adilzainal/IIUM_MBBS_Year4.git
- Download my lecture script https://github.com/adilzainal/IIUM_MBBS_Year4.git
- Learn via recorded lecture – https://classroom.google.com/c/MTI0MzA3NTAyNTky?cjc=b3rsouj
- Synchronous session (2 days)
- Continuous Assessment (10%)

- Not a statistical software
- GNU General public license
- Command line interface and Graphic user interface


Python
SQL
Java
Amazon ML
R
C C++ or C#
Hadoop
Tableau
Apache Spark
SAS
Google
Microsoft Azure
Apache Hive
Scala
MATLAB
SPSS
Tensorflow
Splunk
Apache Pig
Teradata
Stata
Cognos
Scikit Learn
Pytorch
Minitab
Alteryx
Keras
JMP
Caffe
Spotfire
Apache Flink
Databricks
SAP
Apache MXNet
Pentaho
Theano
Mllib
H2O
The More Popular Data Science Software (>=250 Jobs)
0
10000
20000
Number of Data Science Jobs on Indeed.com


SPSS Statistics
R
SAS
Stata
GraphPad Prism
MATLAB
Python
Apache Hadoop
TensorFlow
Java
Minitab
Statistica
SQL
Scikit Learn
JMP
Systat
Keras
C, C++, or C#
Caffe
FORTRAN
Apache Spark
PyTorch
Statgraphics
Mathematica
IBM Watson
Weka
Theano
Data Science Software With >= 1,700 Articles
0
20,000
40,000
60,000
80,000
Number of Scholarly Articles in 2018

JUST A TOOLS

1. Does it run natively on your computer?
2. Does the software provide all the methods you need? If not, how extensible is it?
3. Does its extensibility use its own unique language, or an external one (e.g. Python, R) that is commonly accessible from many packages?
4. Does it fully support the style (programming, or menus and dialog boxes, or workflow diagrams) that you like?
5. Are its visualization options (e.g. static vs. interactive) adequate for your problems?
6. Does it provide output in the form you prefer (e.g. cut & paste into a word processor vs. LaTeX integration)?
7. Does it handle large enough data sets?
8. Do your colleagues use it so you can easily share data and programs?
9. Can you afford it?

PACKAGES

- Graphic display – ggplot2, lattice, plot3D
- Reproducible research – officer, knitr
- Data manipulation – dplyr, reshape2,tidyr
- Machine learning – forecast, lme4
- Bayesian – mcmc
- Spatial – ggmap, gstat
- Meta-analysis - meta

Resources online

- https://www.coursera.org
- https://www.r-bloggers.com
- https://stats.stackexchange.com
- https://stackoverflow.com/questions/tagged/statistics
- https://github.com/tidyverse
- https://ggplot2.tidyverse.org

"Where there is no hope, we must invent it."

Albert Camus