Women's
Data Science Study Group Calgary

Ada Lovelace, considered the first computer programmer.

Short introduction

1. What is your name?

2. What is your background?

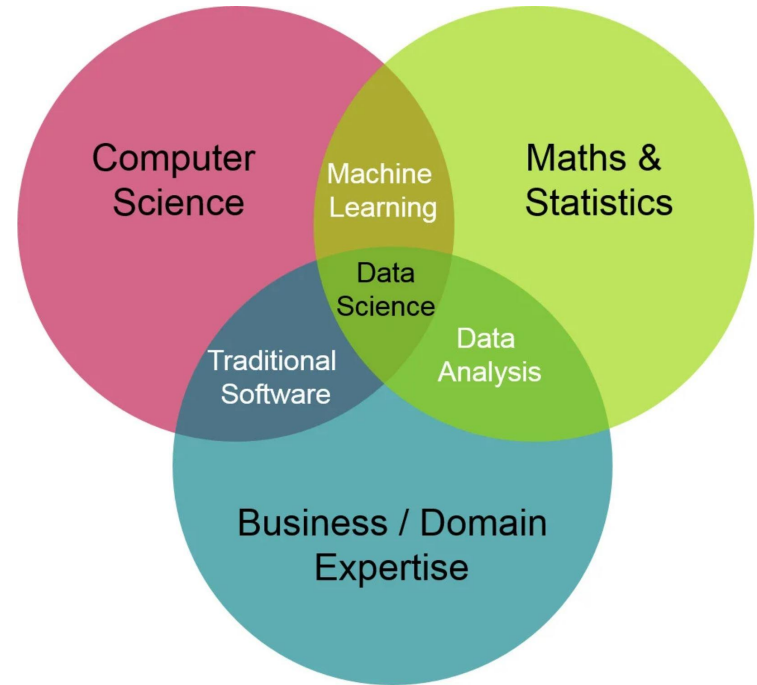3. What would you like to get from participating in this group?

Introduction to R and RStudio
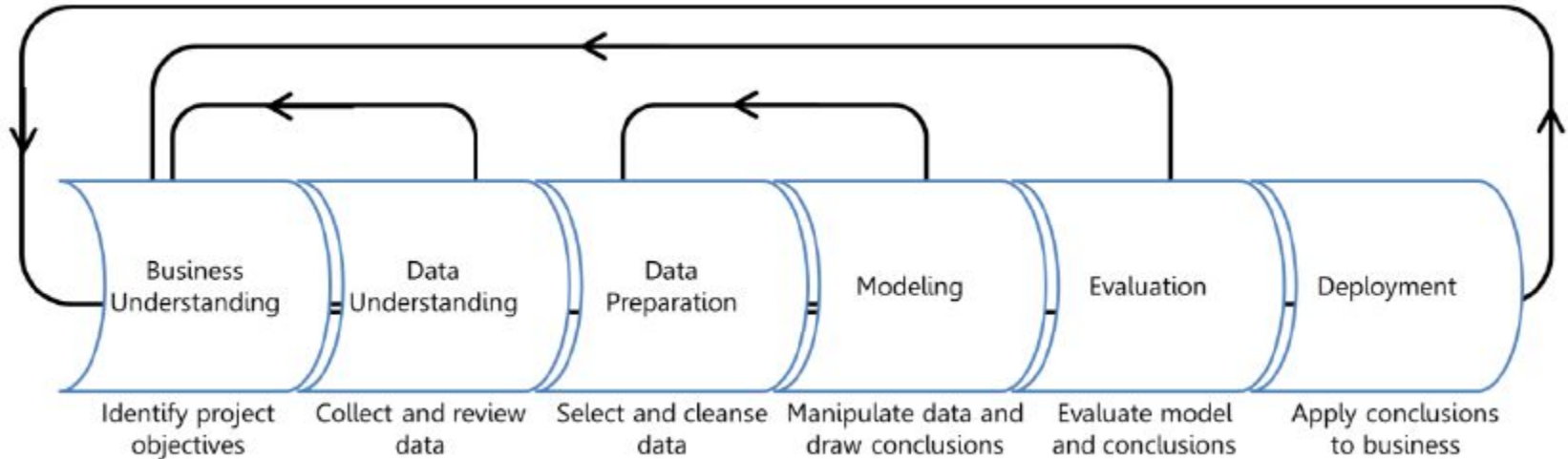
Meeting #1: February 23, 2019

Data Science

Data Science can be defined as a multi-disciplinary subject that encompasses the use of mathematics, statistics, computer science, and business expertise to study and evaluate data.
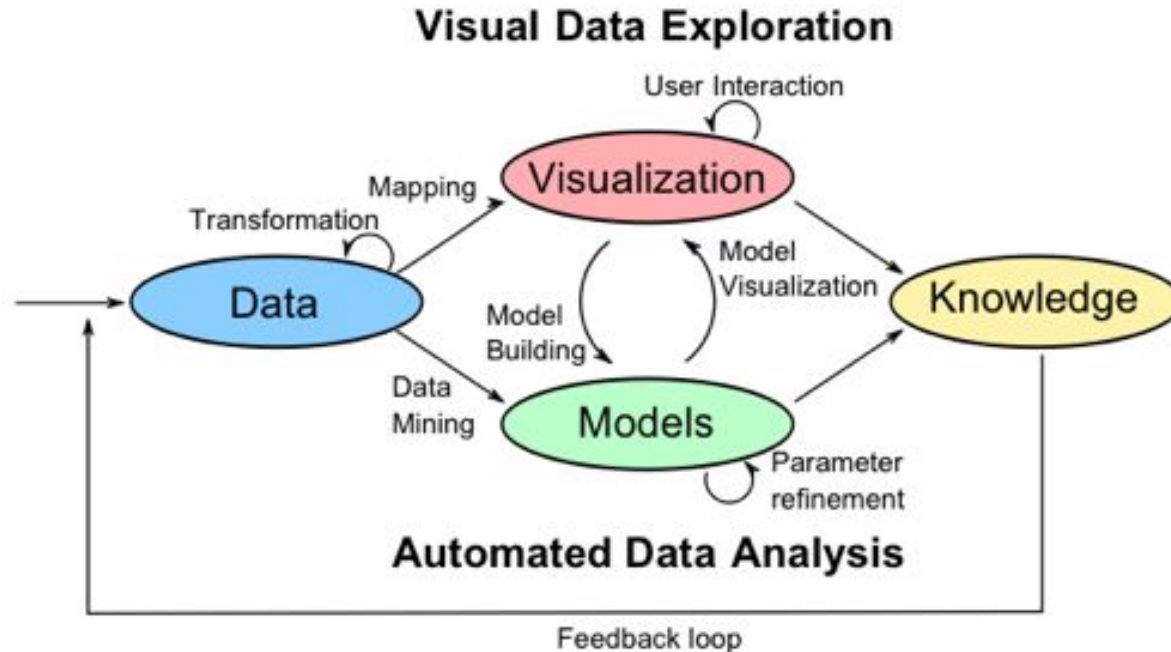
"The key objective of Data Science is to extract valuable information for use in strategic decision making, product development, trend analysis and forecasting."

Data Science Process

Visual Analytics Process

Visual Data Exploration

User Interaction

Visualization

Mapping

Transformation

Data

Model Building

Model Visualization

Knowledge

Data Mining

Models

Parameter refinement

Automated Data Analysis

Feedback loop

http://www.visual-analytics.eu/faq/

Skills Required

	Data Analyst	Machine Learning Engineer	Data Engineer	Data Scientist
Programming Tools	Very important	Very important	Very important	Very important
Data Visualization and Communication	Very important	Somewhat important	Somewhat important	Very important
Data Intuition	Somewhat important	Very important	Somewhat important	Very important
Statistics	Somewhat important	Very important	Somewhat important	Very important
Data Wrangling	Not that important	Not that important	Very important	Very important
Machine Learning	Not that important	Very important	Not that important	Very important
Software Engineering	Not that important	Somewhat important	Very important	Somewhat important
Multivariable Calculus and Linear Algebra	Not that important	Very important	Not that important	Somewhat important

Not that important | Somewhat important | Very important
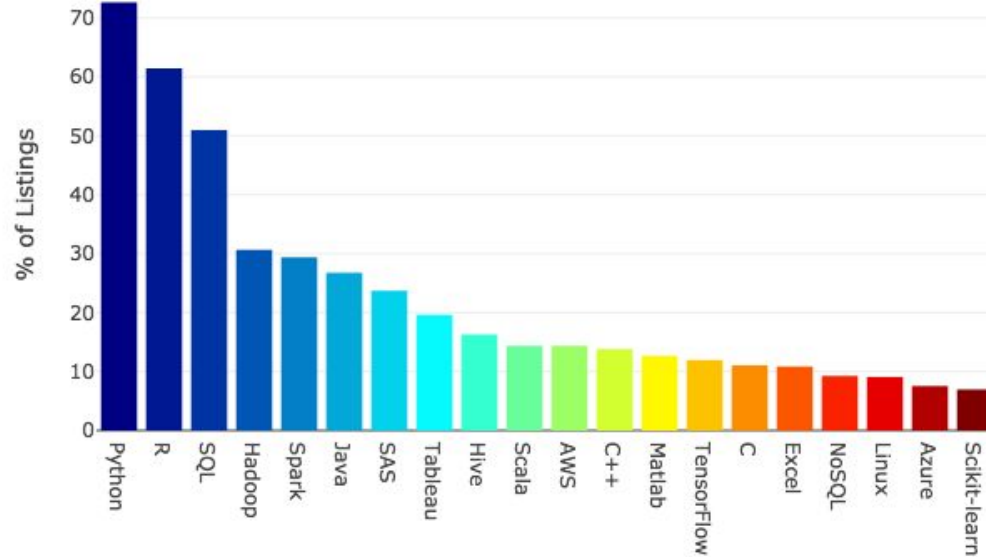
Technology Skills in DS Job Listings in 2018

Top 20 Technology Skills in Data Scientist Job Listings

https://towardsdatascience.com/the-most-in-demand-skills-for-data-scientists-4a4a8db896db

About R

- Free and open source tool
- It has libraries for:
  - Statistics
  - Machine learning
  - Data Science

- Covers a wide range of topics such as econometrics, finance, and time series.

- Best-in-class tools for:
  - Visualization
  - Reporting
  - Interactivity

RStudio is a four pane work-space:

1. Code editor: allows to create and open a file containing R script. The R script is where you keep a record of your work.

2. R console: for typing R commands

3. Workspace tab: shows the list of R objects you created during your R session.
   History tab: shows the history of all previous commands

4. Files tab: shows files in your working directory.
   Plots tab: shows the history of plots you created.
   Packages tab: shows external R packages available on your system. If checked, the package is loaded in R.

http://www.sthda.com/english/wiki/running-rstudio-and-setting-up-your-working-directory-easy-r-programming
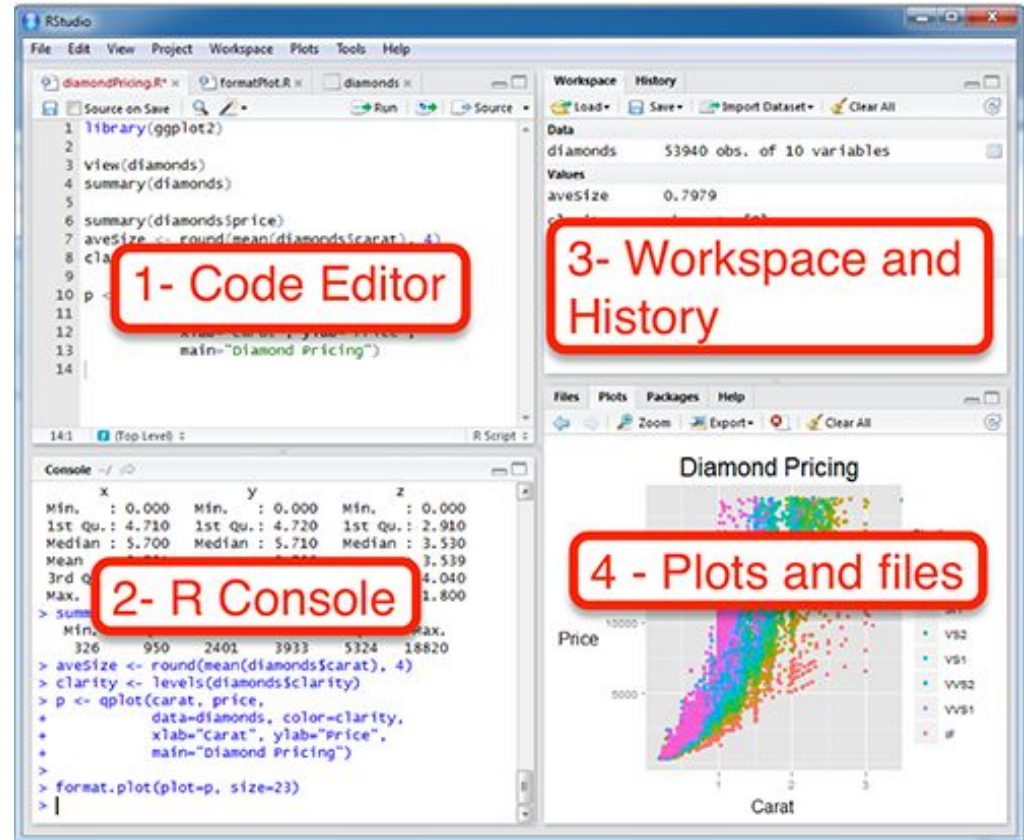
Some R Libraries for Data Analysis

- Data manipulation (`dplyr`, `tidyr`)
- Working with data types:
  - `stringr` for strings
  - `lubridate` for date/datetime
  - `forcats` for categorical/factors
- Visualization (`ggplot2`)
- Programming (`purrr`, `tidyeval`)
- Communication (`Rmarkdown`, `shiny`)
- And many more.

Tidyverse

The **tidyverse** is a collection of R packages developed by RStudio's chief scientist Hadley Wickham.
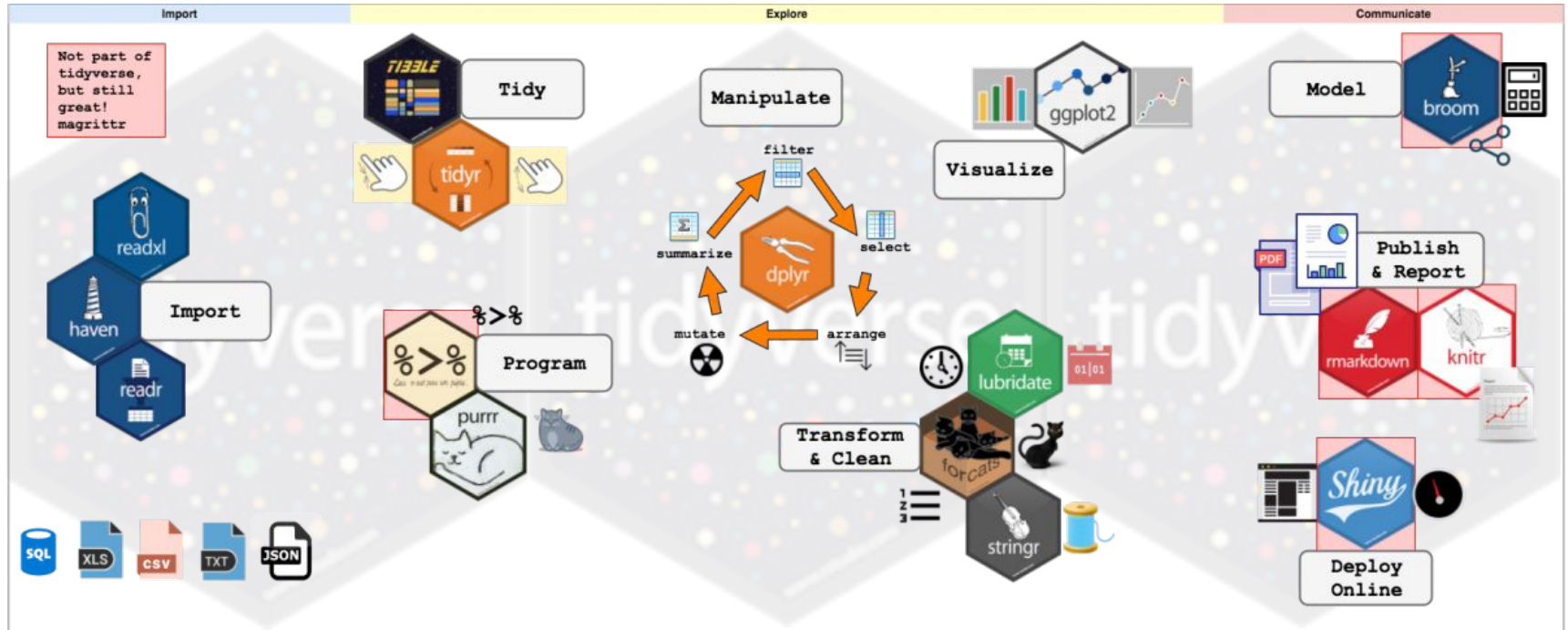
R for Business

Two major advantages of learning R versus every other programming language is that it can produce **business-ready reports** and **machine learning-powered web applications**.

The two capabilities we refer to are

- `rmarkdown` for report generation
- `shiny` for interactive web applications

Rmarkdown

Rmarkdown is a framework for creating reproducible reports or to building blogs, presentations, websites, books, journals, and more.

A few examples of its capability are:

- rmarkdown for generating HTML, Word and PDF reports
- rmarkdown for generating presentations
- flexdashboard for creating web apps via the user-friendly Rmarkdown format.
- blogdown for building blogs and websites
- bookdown for creating online books
- Interactive documents
- Parameterized reports for generating custom reports (e.g. reports for a specific geographic segment, department, or segment of time)

Shiny

Shiny is a framework for creating interactive web applications that are powered by R.

It enables non-data scientists to gain the benefit of data science via interactive decision making tools.

Some examples:
- https://frissdemo.shinyapps.io/FrissDashboard/
- https://gallery.shinyapps.io/087-crandash/
- https://gallery.shinyapps.io/lake_erie_fisheries_stock_assessment_app/
- More Shiny examples at https://www.rstudio.com/products/shiny/shiny-user-showcase/

Now, let's jump off and start learning R