**WEEK 1**

Programing as a data analyst

**Programming** helps:

* Clarify the steps of your analysis
* Save time
* Reproduce and share your work

sudo apt-get install -f

**WEEK 2**

Understand basic programming concepts

**Vector (R)** A group of data elements of the same type stored in a sequence in R

**Pipe (R)** A tool in R for expressing a sequence of multiple operations, represented with “%>%”

Learning about R packages

**Packages include:**

* Reusable R functions
* Documentation about the functions
* Sample datasets
* Tests for checking your code

**CRAN** (Comprehensive R Archive Network) An online archive with R packages, source code, manuals, and documentation

Explore the tidyverse

**Tidyverse (R)** A system of packages in R with a common design philosophy for data manipulation, exploration, and visualization

**Conflicts** happen when packages have functions with the same names as other functions

**For packages** that are an essential part of the workflow for data analysts:

* **ggplot2** Create a variety of data viz by applying different visual properties to the data variables in R
* **dplyr** Offer a consistent set of functions that help you complete some common data manipulation tasks
* **tidyr** A package used for data cleaning to make tidy data
* **readr** Used for importing data

**Factors (R)** Store categorical data in R where the data values are limited and usually based on a finite group like country or year

**Nested** In programming, describes code that performs a particular function and is contained within code that performs a broader function

**WEEK 3**

Explore data in R

**Data frame** A collection of columns

* Columns should be named
* Data stored can be many different types, like numeric, factor
* Each column should contain the same number of data items

**Tibbles**

* Never change the data types of the inputs
* Never change the names of your variables
* Never create row names
* Make printing easier

**Tidy data (R)** A way of standardizing the organization of data within R

**Tidy data standards**

* Variables are organized into columns
* Observations are organized into rows
* Each value must have its own cell

Take a closer look at the data

**Anscombe’s quartet** Four datasets that have nearly identical summary statistics

**WEEK 4**

Create data visualizations in R

**Benefits of ggplot2:**

* Create different types of plots
* Customize the look and feel of plots
* Create high quality visuals
* Combine data manipulation and visualization

**Aesthetic (R)** A visual property of an object in your plot

**Geom (R)** The geometric object used to represent your data

**Facets (R)** Let you display smaller groups, or subsets, of your data

**Labels and annotations (R)** Let you customize your plot

**Mapping (R)** Matching up a specific variable in you dataset with a specific aesthetic

Annotate and save visualizations

**Annotate** To add note to a document or diagram to explain or comment upon it

**WEEK 5**

Create R Markdown documents

**YAML** A language for data that translates it so it’s readable

Understand code chunks and exports

**Delimiter** A character that indicates the beginning or end of a data item