# Git

## Summary

## Coursera

## Johns hopinks specialization xx

### ‘R Programming’

## Tidyverse specialization

### ‘Introduction to the Tidyverse’

#### Terminology

|  |  |  |
| --- | --- | --- |
| **Term** |  |  |
| Dataset | Collection of data |  |
| Observation |  | Eg: pt 1,2,3,4 |
| Variable |  | Eg: id, name, weight, height |
| Type | Same eg patient, different data in other table | (eg address info vs blood level measurements |
| Tidy table | 1 variable per column.  1 observation per row  1 type of data per table  1 column allowing to merge tables |  |
| Rules of tidy tables | * Be consistent * Choose good names for things * Write dates as YYYY-MM-DD * No empty cells * Put just one thing in a cell * Don’t use font color or highlighting as data * Save the data as plain text files |  |
| Relative path |  | specify destination from current workfolder |
| absolute path |  | specify destination from root folder |
| API | Application Programming Interface |  |

## Packages

|  |  |  |
| --- | --- | --- |
| Name | type | notes |
| here | directory tool | similar to getwd, setwd; always starts in .Rproject file path |
| tibble | table format | Table that does less, complains more (no partial matching, doesnt change data types automatically) not a package... |
| readr | Read data into R (eg csv) | import rectangular data (csv, tsf etc), fails if formats are incorrect |
| readxl | Read data from excel into R | especially for .xls or .xlsx files |
| googlesheets4 | read data from googlesheets into R | specifically for google sheets |
| googledrive | read data from google drive into R | goes beyond just google sheets |
| haven | read data from statistical packages | especially SPSS, Stata, SAS |
| jsonlite & xml2 | read semi-structured data | specifically JSON and XML files |
| rvest | read data from internet pages | read from internet pages |
| httr | read from APIs | interact with web API’s |
| dplyr | tidy up data | verbs: mutate, select, filter, summarize, arrange |
| tidyr | tidy up data | verbs:   * gather() * separate() :separate if two variables in one cell * spread() |
| janitor | tidy up data (standardize/fix column names, show table statistics overview, remove empty rows/columns), excel numbers to dates | verbs: see [HERE](https://garthtarr.github.io/meatR/janitor.html#clean_names())   * clean\_names() * tabyl() * remove\_empty() * Crosstabulation() * excel\_numeric\_to\_date() (eg excel 41103 = 2012-07-13) |
| forcats | allows working with categorical data | eg race, sex, etc, all non-number data |
| stringr | allows working with strings | any value written within a pair of single quotes or double quotes in R |
| lubridate | allows working with timezone, time & date |  |
| glue | allows working with interpreted string literals | these are character sequences between double quotes eg “bar” |
| tidytext | allows working with long texts |  |
| purr | allows working with functions and vectors |  |
| skimr | summarizing of (clean) data |  |
| ggplot2 | figure plotting |  |
| ggrepel | ggplot2-adjacent package | repel overlapping text labels for ggplot2 |
| cowplot | ggplot2-adjacent package | polish ggplot2 graphs, make compound figures, aligning etc |
| patchwork | ggplot2-adjacent package | combining multiple plots together |
| kableExtra | table generation | uses ggplot2 syntax |
| gganimate | animation of ggplot2 plots | eg show graph as it changes throughout the year (life-expectancy different continents from 1970-2020) |
| broom & infer | R statistical data results tidying | broom: tidies lm() data (lm() is ‘linear model’, used to create simple regression model  infer: alligns syntax of statistical tests in R for easier comparison |
| tsibble & feast & fable | time series analysis |  |
| fable | forecasting analysis/modeling |  |
|  |  |  |