# R Programming

Some commands in RStudio

* A <- available.packages(); head(rownames(a), 3)
* Install.packages(“slidify”) to install a package
* Install.packages(c(“slidify”, “ggplot2”, “devtools”)) to install packages
* Library(ggplot2) to load packages
* Search() to see all functions
* > data[c(1, 2), ] eerste twee rijen extracten
* > data[47, 1] element op 47de rij, 1ste kolom
* > sum(is.na(data$col)) som van NAs in kolom met naam col

Gemiddelde van kolom na wissen meerdere rijen (met bepaalde andere waardes of Nas)

* x <- data[1] > 31 vector van logicals voor set eerste rijen
* y <- data[4] > 90 vector van logicals voor tweede set rijen
* datagood <- data[x, ] eerste set rijen wissen
* datagood <- datagood[y, ] tweede set rijen wissen
* good <- complete.cases(datagood) vector van logicals voor NAs
* datagood <- data[good, ] rijen met Nas wissen
* mean(datagood[[2]]) bereken gemiddelde 2de colom

## Loop functions

* Flags is een dataframe (194, 30)
* cls\_vector <- sapply(flags, class) returnt vector van classes
* sum(flags$orange) telt aantal landen met oranje in de vlag
* flag\_colors <- flags[, 11:17] extract kolommen met kleuren
* sapply(flag\_colors, sum) vector van aantal vlaggen met kleur
* sapply(flag\_colors, mean) vector van aandeel van kleur
* flag\_shapes <- flags[, 19:23] extract kolommen met vormen
* shape\_mat <- sapply(flag\_shapes, range) matrix (5,2) met vorm en min+max
* vapply(flags, class, character(1)) geeft controle over vorm, veiliger dan sapply
* tapply(flags$animate, flags$landmass, mean) mean van animate per landmass
* tapply(flags$population, flags$red, summary) summary van population voor rood en niet rood

## Simulation

* **rnorm**(aantal, mean, density)
* rnorm(5, 2, 1) [1] 1.49 1.07 3.47 2.03 0.97
* Geeft vector met 5 getallen, mean 0, density 1)
* **rpois**(aantal, mean)
* rpois(10,3) [1] 6 2 3 2 2 2 2 3 2 1
* Geeft een vector met 10 gehele getallen, met als mean 3
* **pnorm**(getal)
* pnorm(1) [1] 0.8413447
* Geeft de kans dat een variabele kleiner is dan getal
* **ppois**(getal, mean)
* ppois(4,2) [1] 0.947347
* Geeft de kans dat een variabele kleiner is dan getal
* set.seed[1] startpositie voor random
* **sample**(set of objects, aantal)
* sample(1:10, 5) [1] 1 8 7 10 6
* selecteert random uit een gegeven set

# Data

Types of question

* Descriptive analysis
* Exploratory analysis: correlations (doesn’t imply causation)
* Inferential analysis: extrapolate small data, involves estimating quantity and uncertainty
* Predictive analysis: using data on x to predict values for y
* Causal analysis: outcomes of variable change
* Mechanistic analysis: identifying variable change leading to other variable change

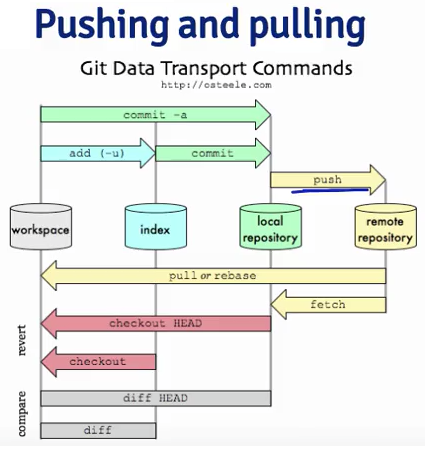
What is data

* Population: set of items you are interested in eg set of websites
* Variables: measurements/characteristics of those items eg height of persons
* Qualitive, quantitative: not ordered vs ordered

Experimental design

Confounding: shoe size -> age -> iq. Account for age, then see results. Check other causal variables

# Git and Github



Pulling:

* mkdir Test-repo
* cd Test-repo
* git init
* git remote add origin https://github.com/Dezwirey/Test-repo.git

or: git clone https://github.com/Dezwirey/Test-repo.git

Adding:

* git add . adds new files
* git add -u updates changed files
* git add -A does both

Committing:

* git commit -m "message"

Pushing:

* git push

Branches:

* git checkout -b branchname creating
* git branch to see what branch
* git checkout master get to master