# Manage Packages

See # Manage R packages in Useful R code.r

# Import Data

dataset <- readr::read\_csv(file.choose())

dataset <- readxl::read\_excel(file.choose())

dataset <- haven::read\_sav(file.choose())

dataset <- rio::import(file.choose()) # many formats

# Data Cleaning

View(dataset)

glimpse(dataset)

str(dataset)

# See column classes  
dataset %>%  
 dplyr::summarise\_all(class) %>%  
 tidyr::gather(variable, class)

# change to factor

dataset %>% mutate(column = as.factor(column)) -> dataset

# order factor levels

dataset %>% mutate(column = fct\_relevel(column,”one”,”two”,”etc”)) -> dataset

# create new variable

dataset %>% mutate (nominal2 = nominal) -> dataset

# reverse code variable

Dataset$ordinal2 <- recode(dataset$ordinal2, first=4, second=3, third=2, fourth=1)

# tidyverse reverse code variable  
dataset %>% mutate(newVariable = recode(originalVariable, “high”=”low”, “medium”=”medium”, “low”=”high”))

# add label to variable

dataset %>% labelled::set\_variable\_labels(ordinal2 = "Variable Label for ordinal2") -> dataset

# add labels to values in variable

dataset %>% labelled::set\_value\_labels(ordinal2 = c(low=1, "medium low"=2, "medium high"=3, high=4)) -> dataset

# change a value

dataset %>% mutate(variable = replace(variable, some\_variable = value, replacementValue) -> dataset

# create a composite variable

dataset %>% group\_by(r = row\_number()) %>% mutate(newVariable = sum(c(as.numeric(var1),as.numeric(var2)), na.rm = FALSE)) -> dataset

# z-scores

dataset %>% group\_by(r = row\_number()) %>% mutate(zscore = scale(variable))

# Data Exploration

attributes(object) # get metadata about object

summarytools::freq(dataset[c("nominal","ordinal")])

summarytools::descr(dataset[c("interval","ratio")])

view(dfSummary(dataset, plain.ascii = FALSE, style = "grid", graph.magnif = 0.75, valid.col = FALSE, tmp.img.dir = "/tmp"))

ggplot(data=dataset, aes(x=nominal)) +

geom\_bar(stat="count")

ggplot(data=dataset, aes(x=ordinal)) +

geom\_bar(aes(y = (..count..)/sum(..count..)))

ggplot(data=dataset, aes(x= "", y = interval)) +

geom\_boxplot()

ggplot(data=dataset, aes(x=ratio)) +

geom\_histogram(binwidth = .2)

ggplot(data=dataset, aes(x=interval, y=ratio)) +

geom\_point()

qplot(dataset$nominal)

qplot(dataset$ordinal)

qplot(dataset$interval)

qplot(dataset$ratio)

qplot(dataset$nominal,dataset$ratio)

# Data Modeling

# Data Export

readr::write\_csv(dataset, path = “folder/fileName.csv”)

readr::write\_rds(dataset, path = "week2data.rds")

haven::write\_sav(dataset, path = "week2data.sav")

openxlsx::write.xlsx(dataset, file = "week2data.xlsx")

rio::export(dataset, file = “filename.ext”, na = “ “) # several formats

rio::convert(“file1.ext”, “file2.ext”)

# Notes

# Format for noting packages and functions

# Can be pasted directly into code.

# This format avoids function masking

# when different packages use the same function name.

packageName::functionName(arguments)

# Useful Shortcuts

Run line of code – Ctrl-Enter

# Resources

https://rstudio.com

Useful sheets - rstudio.com/resources/cheatsheets/

Analysis process cheat sheet - <https://www.business-science.io/r-cheatsheet.html>

R for Researchers: An Introduction: <https://tysonbarrett.com/Rstats/index.html>

R for Data Science: <https://r4ds.had.co.nz/>

Advanced R: <http://adv-r.had.co.nz/Introduction.html>