R for MINNESOTA

Online @ github.com/tidy-mn/cheatsheets

Use this reference to read data, clean column names, make plots, process dates, filter and summarize data, join tables, and more.

PROJECT LAUNCH

New Project

`File` > `New Project` > `New Directory`

New R script

`File` > `New File` > `R Script`

Where am I?

```
# Show current working directory
getwd()
# Set new working directory
setwd("C:/my-data-folder")
```

Install new packages

```
install.packages("readr")
library(readr)
```

STORE VALUES

```
# Use the left-arrow
age <- 7.2

# Text goes in quotes
porg <- "Sunshine"

# Multiple values go inside c()
droids <- c("BB8", "R2D2", "C-3PO")

# Copy an object
my_droids <- droids

# Avoid numbers, spaces, & symbols
3-droids* <- "error_invalid_name"
```

READ DATA

Text files *(.csv, .txt, .tab)*

```
library(readr)
porgs <- read_csv("txt_file.csv")
```

Excel files *(.xlsx, .xls)*

```
library(readxl)
porgs <- read_excel("Excel_file.xlsx")
```

DESCRIBE DATA

```
library(dplyr)
nrow(porgs)
names(porgs)
summary(porgs)
glimpse(porgs)
class(porgs)
# View unique column values
distinct(porgs, age)
## 5 6 11 12 3
```

CLEAN NAMES

```
# Simplify all column names
library(janitor)
porgs <- clean_names(porgs)

# Assign new names manually
library(dplyr)
# Put new name on left: new_name = oldName
rename(porgs, mass_kg = massKG)
```

ADD COLUMNS

```
library(dplyr)

# Add home planet column
mutate(porgs,
       planet = "Earth")
# Add new calculated column
mutate(porgs,
       growth = height / age)
```

FILTER

```
library(dplyr)
# Keep only Porgs older than 3
filter(porgs, age > 3)
# Keep rows with name Jumpity
filter(porgs, name == "Jumpity")
# Keep Porgs named Jumpity OR Chicken
filter(porgs, name %in% c("Jumpity", "Chicken"))
```

SUMMARIZE

```
library(dplyr)
# Summarize the age for the entire table
summarize(porgs, avg_age = mean(age))
# Summarize the age for each color group
group_by(porgs, color) %>%
    summarize(avg_age = mean(age))
```

COMPARISONS

Symbol	Comparison
`>`	greater than
`>=`	greater than or equal to
`<`	less than
`<=`	less than or equal to
`==`	equal to
`!=`	NOT equal to
`%in%`	is value X in list: `X %in% c(1,3,5)`
`is.na(...)`	is the value missing?

PLOTS

Scatterplot

```
library(ggplot2)

ggplot(porgs, aes(x = name, y = age)) +
  geom_point(size = 8, color = "hotpink")
```

Add titles & lines

```
ggplot(porgs, aes(x = name, y = age)) +
  geom_point(size = 8, color = "hotpink") +
  geom_hline(yintercept = 5,
             linetype   = "dashed") +
  labs(title    = "Porgs",
       subtitle = "Sampled on planet Ahch-To",
       caption  = "data from New Republic")
```

Facet by group

```
ggplot(porgs, aes(x = name, y = age)) +
  geom_point(aes(color = color), size = 8) +
  facet_wrap(~color) +
  scale_color_manual(values = c("gray", "yellow")) +
  theme_dark()
```