https://github.com/SBGalvin/Psyc-Wellbeing-2022

Go to the above address for support materials

I used to be scared of using ...

Open Science Tools

...in psychological research

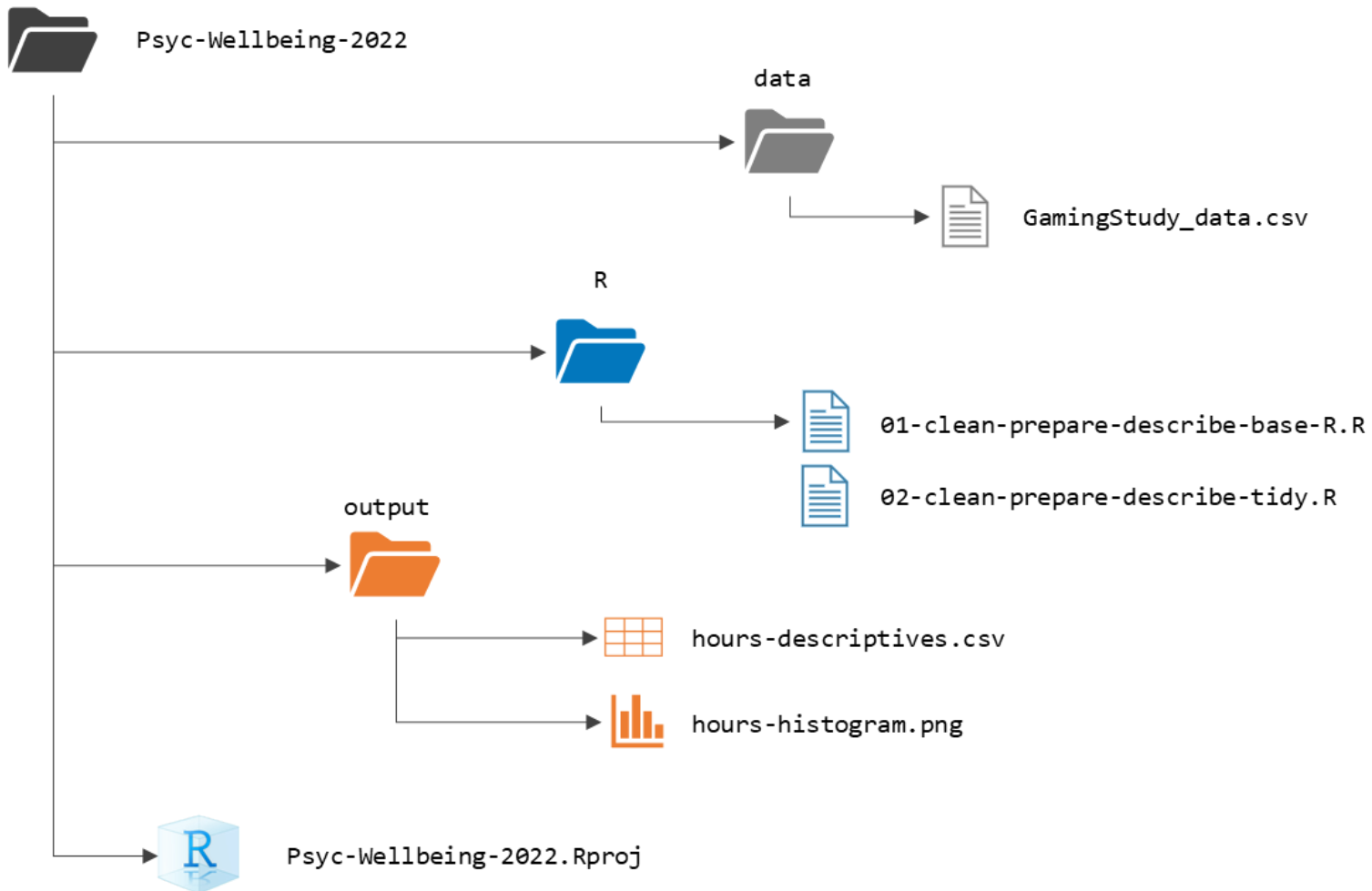
SB Galvin
School of Psychology
DCU
Ollscoil Chathair
Bhaile Átha Cliath
Dublin City University

Session Repository

https://github.com/SBGalvin/Psyc-Wellbeing-2022
Git - Version control software
GitHub - Internet hosting for software development and version control using Git (owned by Microsoft)

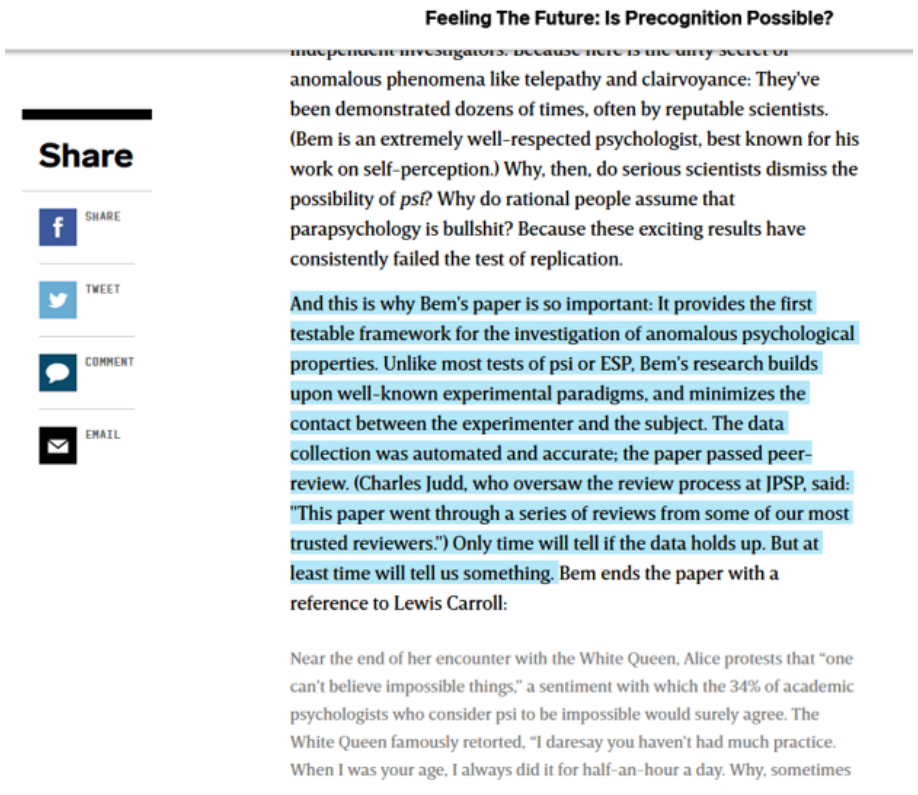
Project Structure

R projects help to create and maintain a workspace for a project. Projects can be simply structured.



The Replication Crisis

Highly exciting, novel findings...



Feeling The Future: Is Precognition Possible?

Share

SHARE

TWEET

COMMENT

EMAIL

anomalous phenomena like telepathy and clairvoyance: They've been demonstrated dozens of times, often by reputable scientists. (Bem is an extremely well-respected psychologist, best known for his work on self-perception.) Why, then, do serious scientists dismiss the possibility of *psi*? Why do rational people assume that parapsychology is bullshit? Because these exciting results have consistently failed the test of replication.

And this is why Bem's paper is so important: It provides the first testable framework for the investigation of anomalous psychological properties. Unlike most tests of psi or ESP, Bem's research builds upon well-known experimental paradigms, and minimizes the contact between the experimenter and the subject. The data collection was automated and accurate; the paper passed peer-review. (Charles Judd, who oversaw the review process at JPSP, said: "This paper went through a series of reviews from some of our most trusted reviewers.") Only time will tell if the data holds up. But at least time will tell us something. Bem ends the paper with a reference to Lewis Carroll:

Near the end of her encounter with the White Queen, Alice protests that "one can't believe impossible things," a sentiment with which the 34% of academic psychologists who consider psi to be impossible would surely agree. The White Queen famously retorted, "I daresay you haven't had much practice. When I was your age, I always did it for half-an-hour a day. Why, sometimes

Image source: https://www.wired.com/2010/11/feeling-the-future-is-precognition-possible/

Poor statistical practice...

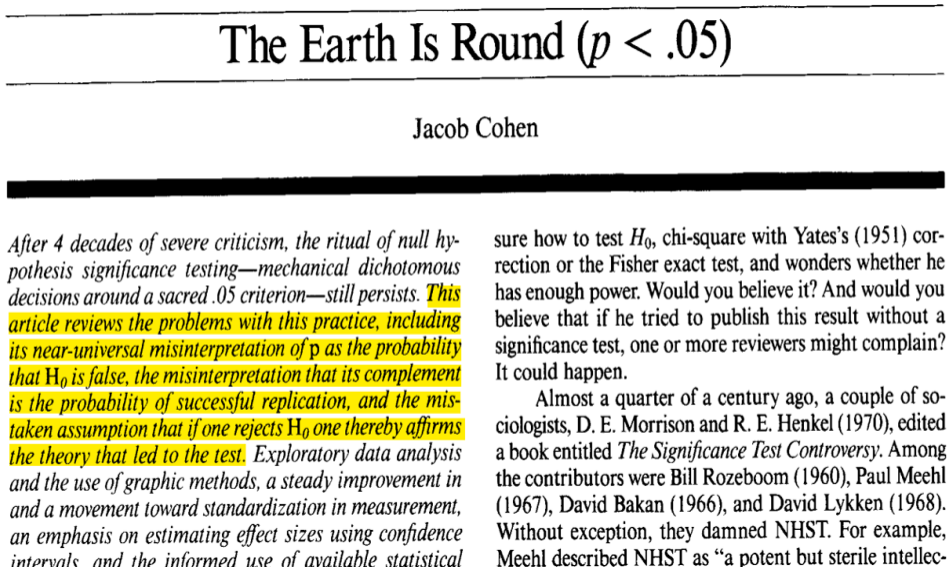
- Critical data are likely unavailable (File drawer problem)
  - (Francis, 2012, Schimmack, 2012)
- Poor methodological rigour (likely p-hacking)
  - (LeBel & Peters, 2011)

Other Examples:

- "Himmicanes vs Hurricanes"
  - Jung, Shavitt, Viswanathan, and Hilbe (2014)
- Power Posing
  - Carney, Cuddy and Yap (2010)
- Positivity/ Losada Ratio
  - Fredrickson and Losada (2005)
- Anything by Brian Wansink

P values I

Historically, p-values have been the primary tool for deciding whether piece of research found something or not.



The Earth Is Round ($p < .05$)

Jacob Cohen

After 4 decades of severe criticism, the ritual of null hypothesis significance testing—mechanical dichotomous decisions around a sacred .05 criterion—still persists. This article reviews the problems with this practice, including its near-universal misinterpretation of p *as the probability that* H_0 *is false, the misinterpretation that its complement is the probability of successful replication, and the mistaken assumption that if one rejects* H_0 *one thereby affirms the theory that led to the test. Exploratory data analysis and the use of graphic methods, a steady improvement in and a movement toward standardization in measurement, an emphasis on estimating effect sizes using confidence intervals, and the informed use of available statistical*

sure how to test H_0, chi-square with Yates's (1951) correction or the Fisher exact test, and wonders whether he has enough power. Would you believe it? And would you believe that if he tried to publish this result without a significance test, one or more reviewers might complain? It could happen.

Almost a quarter of a century ago, a couple of sociologists, D. E. Morrison and R. E. Henkel (1970), edited a book entitled *The Significance Test Controversy*. Among the contributors were Bill Rozeboom (1960), Paul Meehl (1967), David Bakan (1966), and David Lykken (1968). Without exception, they damned NHST. For example, Meehl described NHST as "a potent but sterile intellec-

What a p values is:

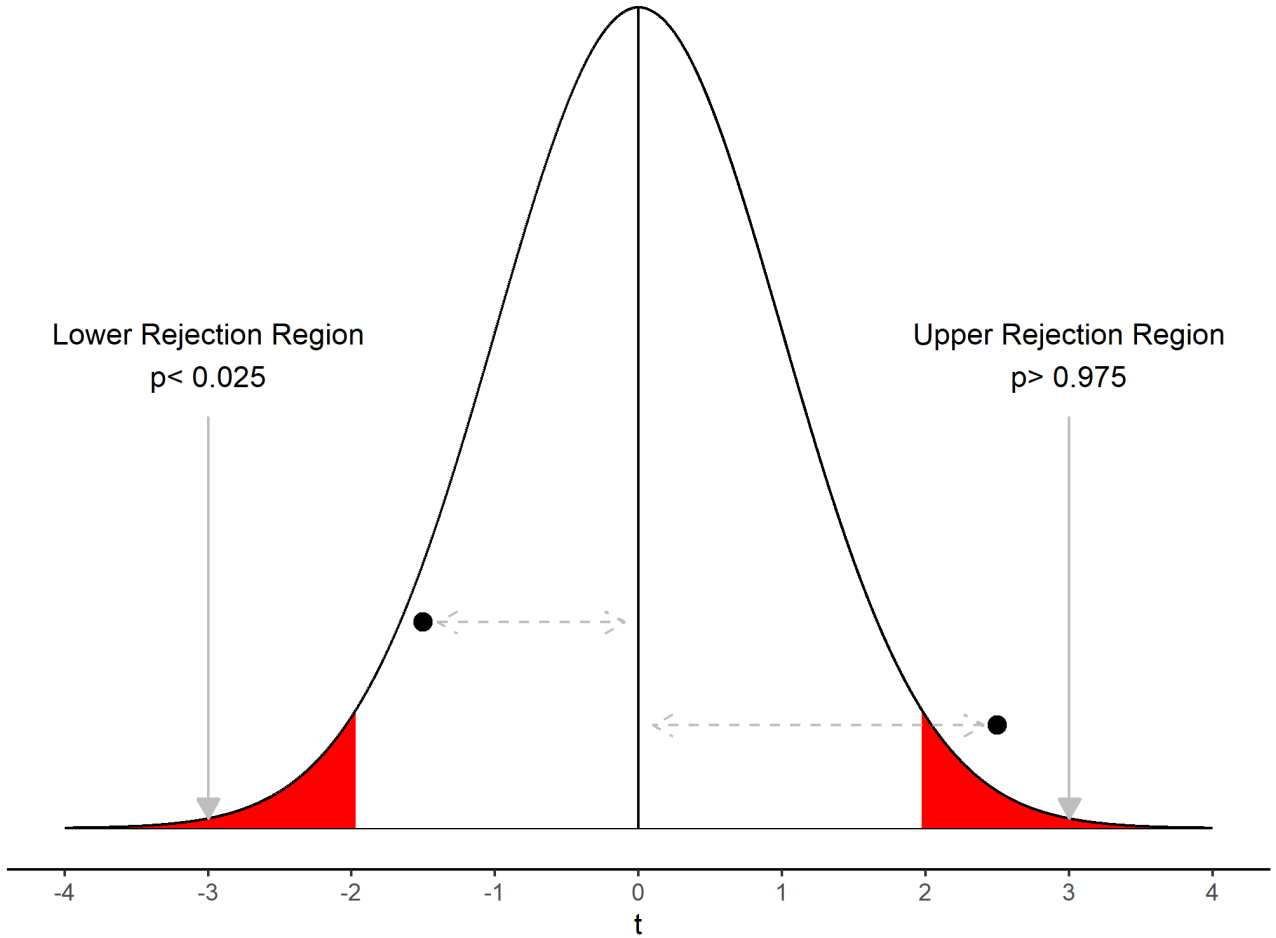
$$P\ value = P(D|H_0)$$

What it tells you:

- *Assuming the null hypothesis to be true, how likely is it to observe this OR more extreme data.*

P values II

In NHST a Null Hypothesis is represented by a null t distribution; a distribution of test-statistics *assuming no TRUE difference*. If a high (or low) p value is obtained, then **WE INFER** observed data is unlikely to have been produced **under the assumption** the Null distribution is true.



Replication Crisis: Focal Points and Solutions

A general principle of transparency is a major thematic undercurrent in an effort to identify research behaviors which influence type I error rates, distinct from fraudulent behaviours, falsifying/fabricating data.

Questionable Research Practices.

- P-Hacking
- HARKing
- Selective outlier deletion
- Selective reporting of (dependent) variables
- Optional Stopping
- Failure to disclose experimental conditions
- Researcher degrees of freedom

Solutions

- General Transparency
- Reproducible work
- Open Data, Open Materials, Open (Analysis) Code
- Pre-registration
- Meta Analyses

Data Analysis

Data analysis is complex, and doesn't start with a statistical test.

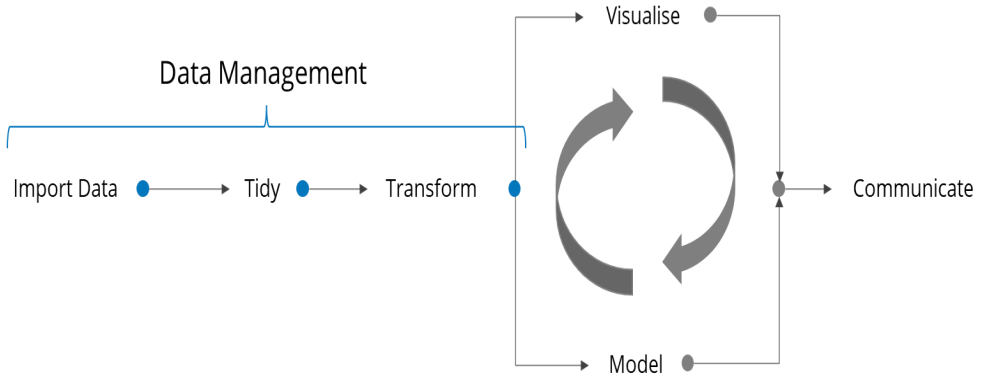


Image source: Grolemund & Wickham (2017) https://r4ds.had.co.nz/explore-intro.html

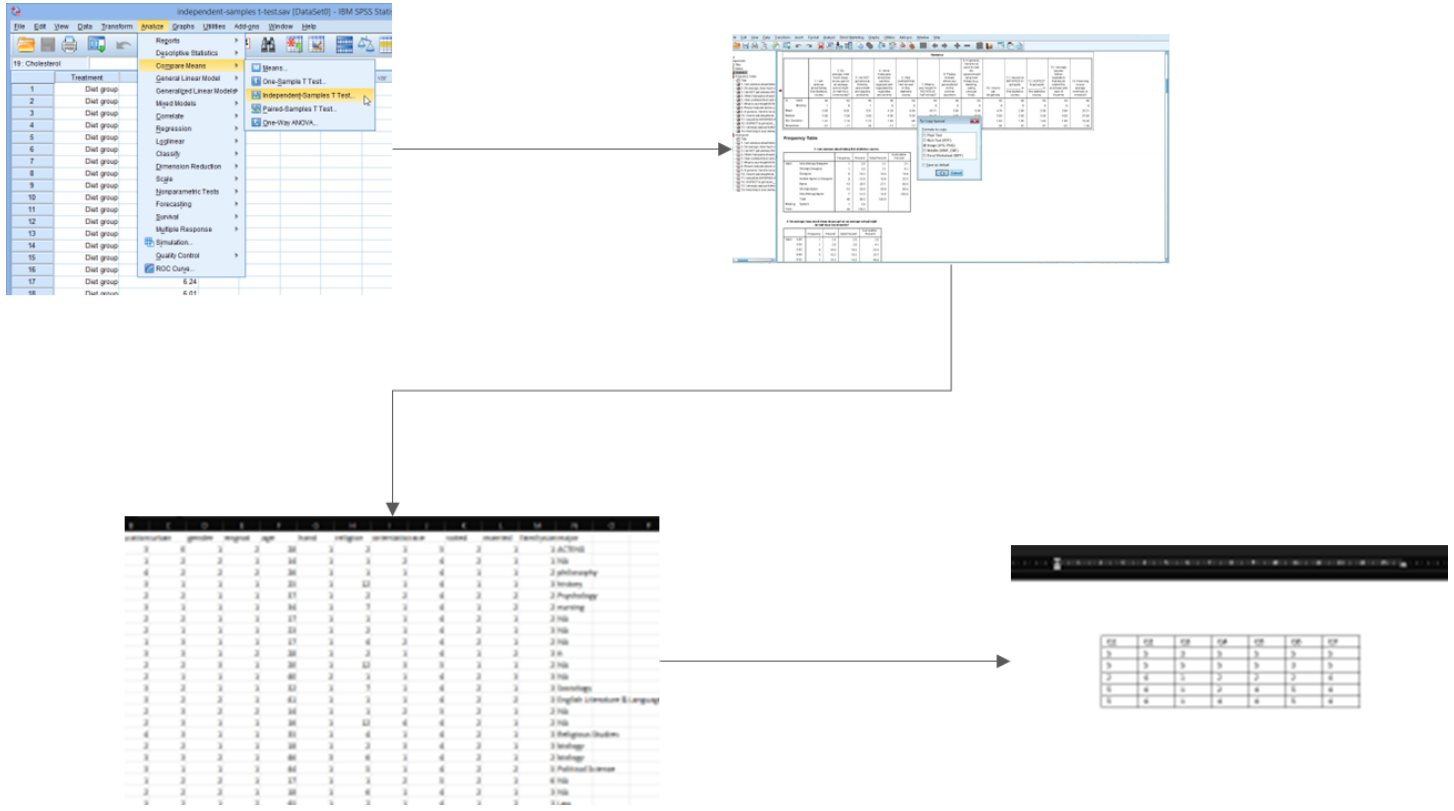
The researcher is not neutral: At each stage of data analysis there are a garden of choices to make!
Treatment of extreme values; Variable Selection and Inclusion; Statistical model choice

Many Analysts, One Dataset: Making Transparent How Variations in Analytic Choices Affect Results
Silberzahn, et al., 2018

- Twenty-nine teams involving 61 analysts used the same data set to address the same research question: whether soccer referees are more likely to give red cards to dark-skin-toned players than to light-skin-toned players.
- Twenty teams (**69%**) found a statistically significant positive effect, and 9 teams (**31%**) did not observe a significant relationship.
- The **29** different analyses used **21** unique combinations of covariates.

Subjective Choice

When we analyse data there are lots of moving parts that may get missed when bouncing between multiple pieces of software.



Using R

Terms, Functions, Scripting

RStudio IDE

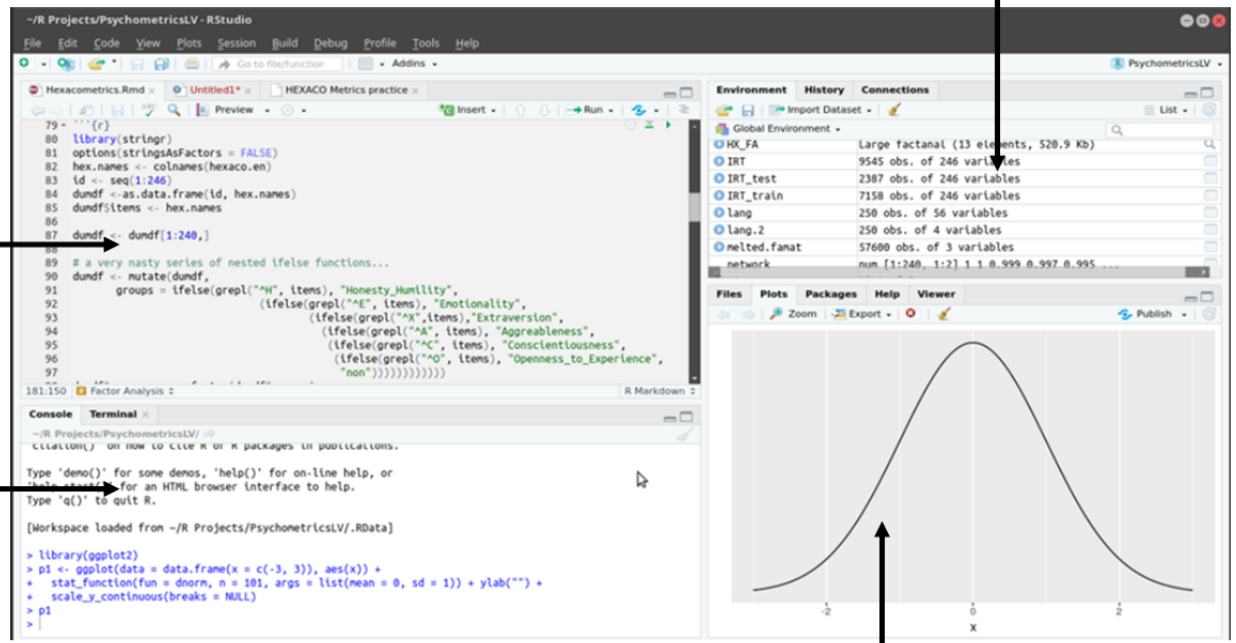
Rstudio is an Integrated Development Environmentthat provides facilities for writing R code, producing/viewing data and graphics.

1 | Source pane

(R scripts etc.)

2 | Environment pane

Objects
variables, dataframes



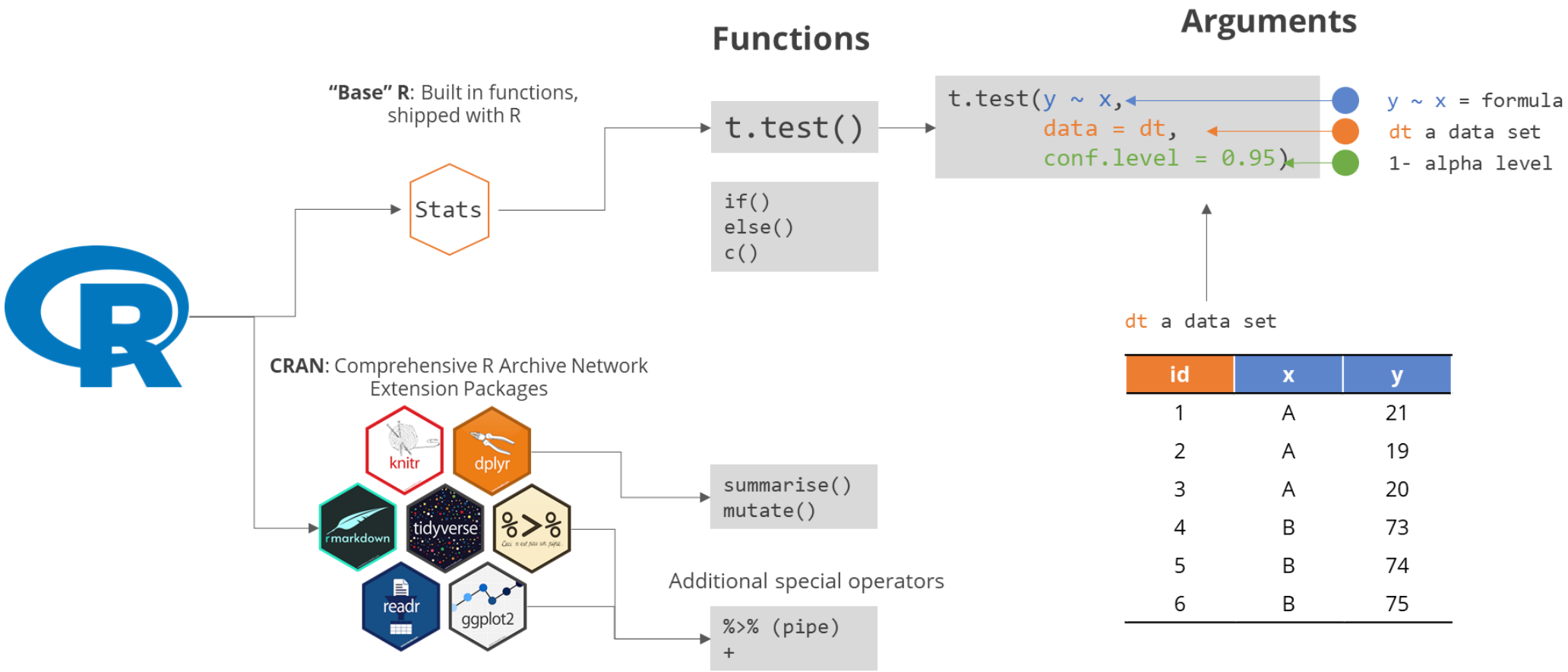
4 | Console pane

R code
Error messages
Output

3 | Viewer pane

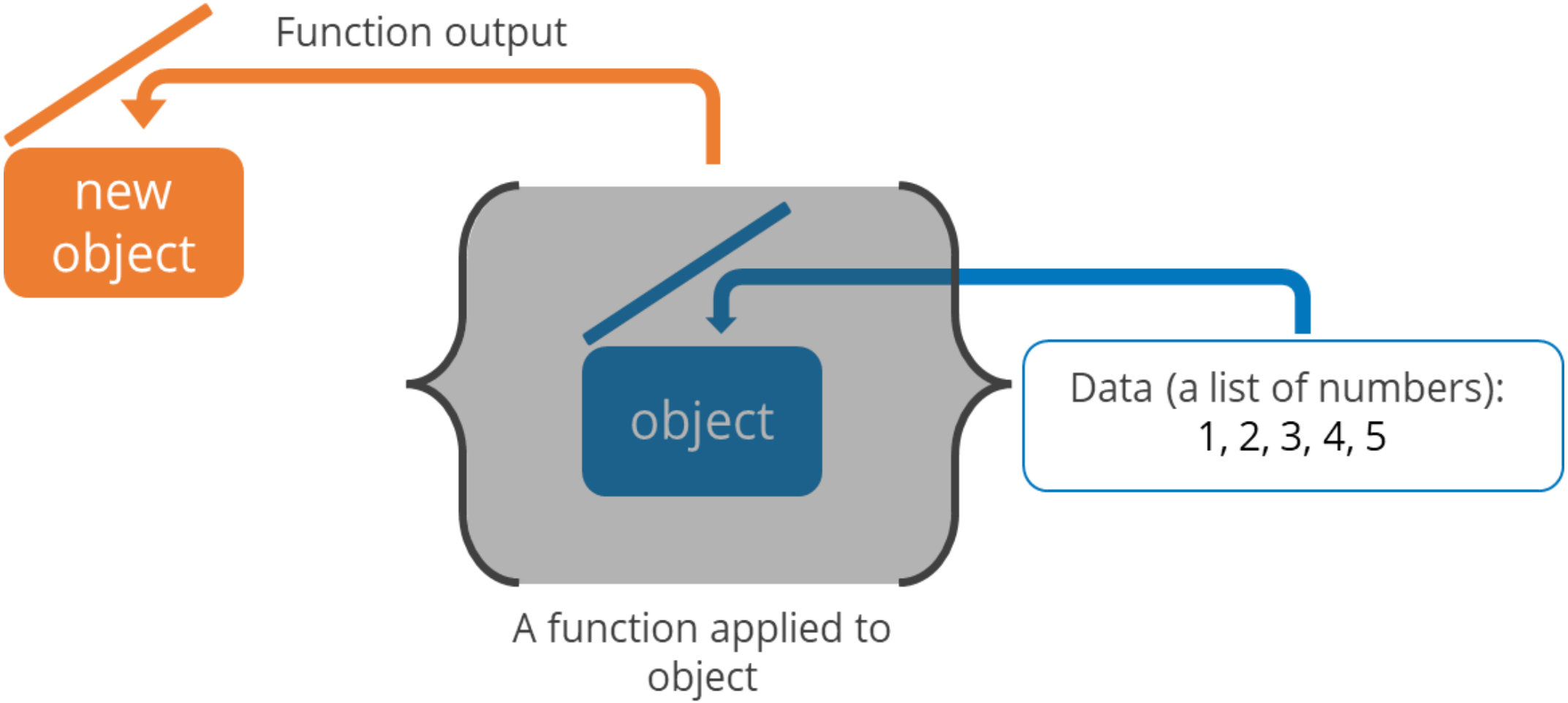
Plots
File viewer
Function help docs

The R Eco-System



id	x	y
1	A	21
2	A	19
3	A	20
4	B	73
5	B	74
6	B	75

R Code


Function output
new object
object
Data (a list of numbers):
1, 2, 3, 4, 5
A function applied to object

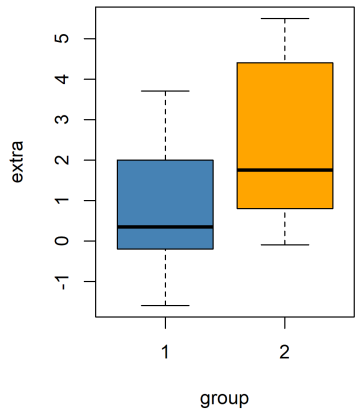
R Examples

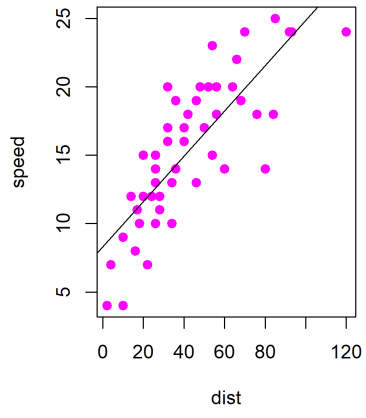
(i) Simple (ish) syntax; (ii) fast to produce statistical graphics; (iii) Massive extensibility; (iv) Reproducible (scripting); (v) Extensive documentation and examples.

Correlations and T-tests

```
cor.test(~ dist + speed, data = cars)    # statistic
plot(~ dist + speed, data = cars)        # scatterplot
```

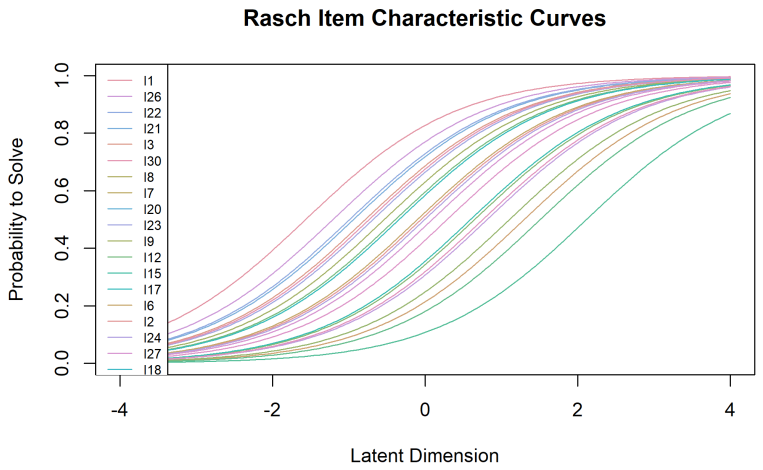
```
# two sample t-test
t.test(extra ~ group, data = sleep)      # statistic
boxplot(extra ~ group, data = sleep)     # box plot

# one-sample test
t.test(extra ~ 1, data = sleep)          # statistic
```





Rasch Models

```
library(eRm)
RaschRes <- RM(raschdat1, se = FALSE, sum0 = TRUE) #  Rasch Model
LRtest(RaschRes, splitcr = "median")               # Likelihood Ratio Test

# Rasch ICC plot
plotjointICC(RaschRes,
             main = "Rasch Item Characteristic Curves",
             legpos = FALSE)
```



Confirmatory Factor Analysis

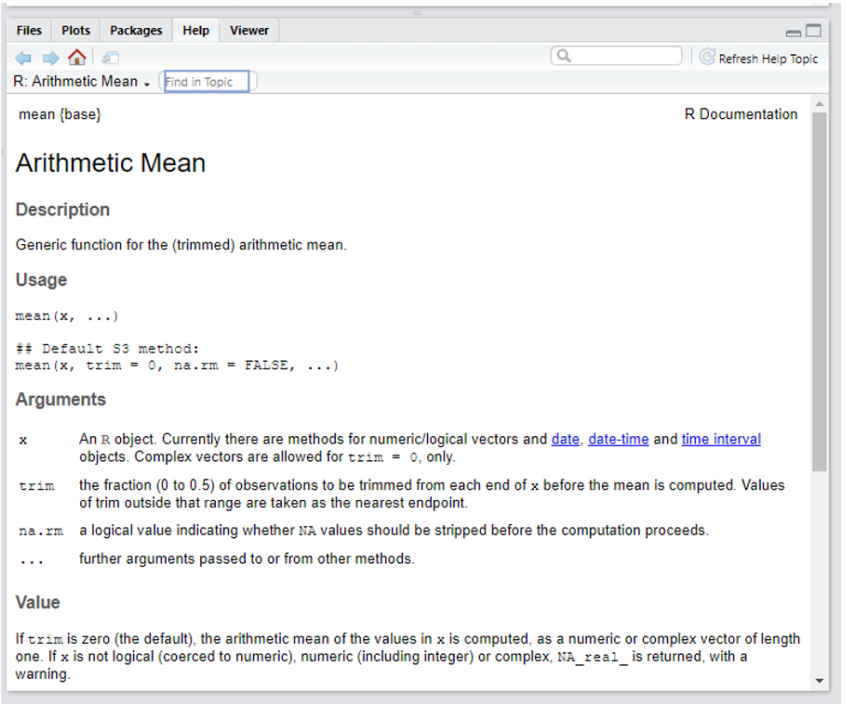
```
library(lavaan)
## The famous Holzinger and Swineford (1939) example
HS.model <- ' visual  =~ x1 + x2 + x3
              textual =~ x4 + x5 + x6
              speed   =~ x7 + x8 + x9 '

fit <- cfa(HS.model, data = HolzingerSwineford1939) # CFA data to model
summary(fit, fit.measures = TRUE)                   # CFA fit values
```

Getting Help

Getting Help

1. Read the documents:
    - R package documents
    - Package repositories
2. Read the error message and try to figure it out
3. Go to Google and copy and paste the error message
4. Go to stack exchange and copy and paste the error message

```
?mean # get help for mean() function
```

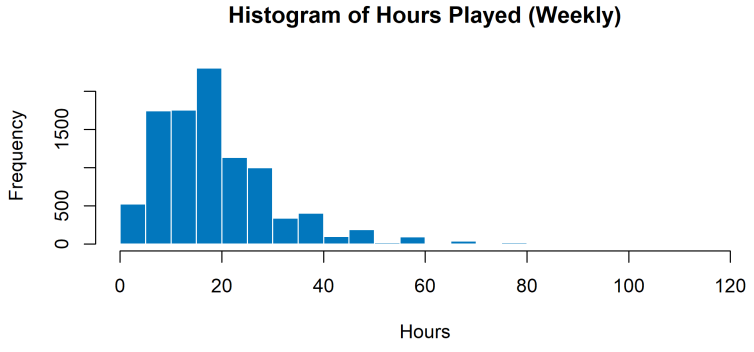


Document help page for the mean function from Rstudio help tab

Writing R Code

See R script R/01-clean-prepare-describe-base-R.R

```r
# R/01-clean-prepare-describe-base-R.R

# 00) Notes  ------------------------------------------------------------
# This script reads in data and prepares it for analysis

# 01) options and settings ----------------------------------------------
options(scipen = 999)
set.seed(42)


# 02) Read in data ------------------------------------------------------
GamingStudyData <- read.csv('data/GamingStudy_data.csv')


# 03) Remove Missing values  --------------------------------------------
# GD_1 remove missing values for Hours
# GD_2 remove hours greater than 168 or less than 0
# GD_3 remove missing values for SPIN
# GD_4 keep only league of legends players
# GD_5 keep non-professional players

GD_1 <- GamingStudyData[!is.na(GamingStudyData$Hours), ]
GD_2 <- GD_1[GD_1$Hours > 0 & GD_1$Hours < 168, ]
GD_3 <- GD_2[!is.na(GD_2$SPIN_T), ]
GD_4 <- GD_3[GD_3$Game     == "League of Legends", ]
GD_5 <- GD_4[GD_4$earnings == "I play for fun", ]

GamingStudyData_clean <- GD_5         # keep last for clean data

rm(GD_1, GD_2, GD_3, GD_4, GD_5)      # remove objects


# 04) Store Data set ----------------------------------------------------
write.csv(GamingStudyData_clean, 'data/GamingStudyData_clean-base-R.csv')

# 05) Histogram and summary of Hours  -----------------------------------

hist(GamingStudyData_clean$Hours,
     main   = "Histogram of Hours Played (Weekly)",
     xlab   = "Hours",
     col    = "#0277BD",
     border = 'white',
     breaks = seq(0, 120, by = 5))

# summary
Hours_summary <- as.data.frame(psych::describe(GamingStudyData_clean$Hours) )
Hours_summary[, c("n", "mean", "median", "sd", "se", "min", "max")]
```



n	mean	median	sd	se	min	max
9702	20.5	20	12.4	0.13	1	120

Simple Rules:

1. Use comments (#) to signpost what the code is doing
2. Space out code
3. Name an R script appropriately/descriptively
4. Check for typos (most common reason code doesn't run)
5. Use descriptive names for objects

Reproducibility in R

Examples with the tidyverse

Tidyverse Packages

Tidyverse is a collection of packages based on the same styling under the **Grammar of Data Manipulation**.
A function is a **verb** it does something to arguments

- web page: https://www.tidyverse.org/
- style guide: https://style.tidyverse.org/index.html

readr: for reading/writing data files

dplyr: for managing/ manipulating data

magrittr: for making code more readable

ggplot2: for data visualisation

haven: for reading/writing data files from other software packages (Stata, SPSS)

Dplyr Syntax

dplyr is designed to be human readable syntax - the functions are named descriptively. When used in combination with the %>% (pipe) operator, the workspace can be kept free of unnecessary objects. %>% pushes data through multiple functions to create an object.

- `mutate()` alters an existing variable or creates a new variable
- `filter()` allows you to filter data depending on a logical condition
- `select()` allows you to select columns in a data set
- `summarise()` allows you to make custom summaries by placing functions inside `summarise()`
- `group_by()` and `ungroup()` groups a table (by specified variable) and ungroups a table.
- `arrange()` sorts a table by a named variable, `desc()` makes this order descending.
- `top_n()` takes the first n rows of a sorted table

```
GamingStudyData_clean <- GamingStudyData %>%
                         filter(!is.na(Hours)) %>%
                         filter(Hours > 0 & Hours < 168) %>%
                         filter(!is.na(SPIN_T)) %>%
                         filter(Game == "League of Legends") %>%
                         filter(earnings == "I play for fun") %>%
  mutate(Game = tolower(Game)) %>%
  select(-earnings)
```

```
GamingStudyData_clean %>%

  group_by(Residence) %>%
  summarise(N = n()) %>%
  ungroup() %>%

  arrange(desc(N)) %>%
  top_n(n = 5)
```

Residence	N
USA	3382
Germany	1004
UK	743
Canada	723
Netherlands	368

Read and Clean

See script R/02-clean-prepare-describe-tidy.R


```
# 02-clean-prepare-describe-tidy.R

# 00) Notes  ------------------------------------------------------------
# This script reads in data and prepares it for analysis
# In tidyverse! (dplyr, ggplot2 packages)

# 01) options and settings ----------------------------------------------
library(tidyverse)
options(scipen = 999)
set.seed(42)


# 02) Read in data ------------------------------------------------------
GamingStudyData <- read_csv('data/GamingStudy_data.csv')


# 03) Remove Missing values  --------------------------------------------

# (1) remove missing values for Hours
# (2) remove hours greater than 168 or less than 0
# (3) remove missing values for SPIN
# (4) keep only league of legends players
# (5) keep non-professional players

GamingStudyData_clean <- GamingStudyData_clean %>%
                         filter(!is.na(Hours)) %>%
                         filter(Hours > 0 & Hours < 168) %>%
                         filter(!is.na(SPIN_T)) %>%
                         filter(Game == "League of Legends") %>%
                         filter(earnings == "I play for fun")

# 04) Store Data set ----------------------------------------------------
write_csv(GamingStudyData_clean, 'data/GamingStudyData_clean-base-R.csv')
```

1. Less objects are created (less messy)
2. Functions make semantic sense
3. The pipe (%>%) helps to space out things

- We can reproduce our cleaning procedure (filter functions)
- Store the resulting clean data as a separate file

Summarise Data

dplyr functions are written to be human readable. The pipe operator (%>%) pushes data forward through multiple functions to help avoid

Creating a manual summary data frame

```
GamingStudyData_clean %>%
  # summarise a single variable
  summarise(N    = nrow(.),
            Mean = mean(Hours),
            SD   = sd(Hours),
            SE   = sd(Hours)/sqrt(nrow(.)),
            Min  = min(Hours),
            Max  = max(Hours)) %>%

  # round numeric values
  mutate_if(is.numeric, round, 2)
```

N	Mean	SD	SE	Min	Max
9702	20.5	12.4	0.13	1	120

Creating a grouped summary table is relatively easy

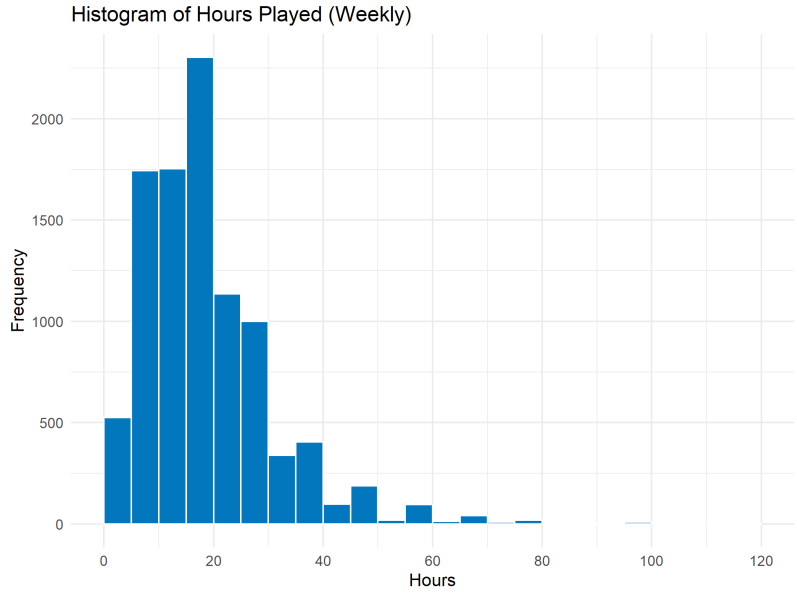
```
GamingStudyData_clean %>%

  # group data
  group_by(Gender) %>%

  # summarise a single variable
  summarise(N    = n(),
            Mean = mean(Hours),
            SD   = sd(Hours),
            SE   = sd(Hours)/sqrt(n()),
            Min  = min(Hours),
            Max  = max(Hours)) %>%
  ungroup() %>%

  # round numeric values
  mutate_if(is.numeric, round, 2)
```

Gender	N	Mean	SD	SE	Min	Max
Female	540	17.96	11.71	0.12	1	75
Male	9130	20.63	12.37	0.13	1	120
Other	32	27.56	21.95	0.22	5	100

Visualise Data

GGplot2 builds up a plot in layers; (1) create a mapping of data to aesthetics; (2) add a geometry to render aesthetics/data; (3) add labels; (4) select thematic settings.

```
GamingStudyData_clean %>%

  # Map aesthetics
  ggplot(mapping = aes(x = Hours))+

  # Use histogram geometry
  geom_histogram(fill = "#0277BD",
                 colour = 'white',
                 breaks = seq(0, 120, by = 5))+

  # Plot/axis labels
  xlab("Hours")+
  ylab("Frequency")+
  ggtitle("Histogram of Hours Played (Weekly)")+

  # scale definitions
  scale_x_continuous(breaks = seq(0, 120, by = 20))+

  # Thematic settings
  theme_minimal()
```



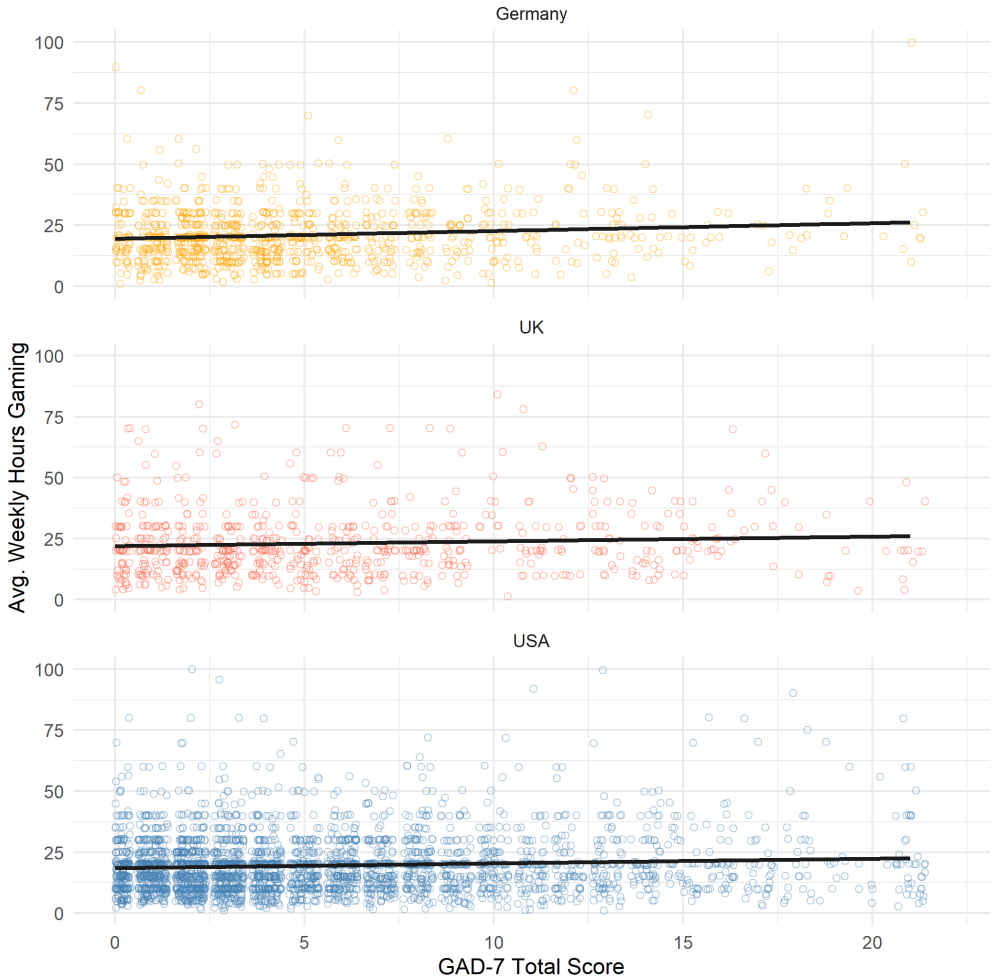
Inferential Statistics

See R script R/03-additional-R-scripts/04-00-Correlation-by-country.R

```
US_hours_GAD_cor <- with(GamingData_clean_US,
                         cor.test(GAD_T, Hours)) %>% tidy()


DE_hours_GAD_cor <- with(GamingData_clean_DE,
                         cor.test(GAD_T, Hours)) %>% tidy()


UK_hours_GAD_cor <- with(GamingData_clean_UK,
                         cor.test(GAD_T, Hours)) %>% tidy()

# table
rbind(US_hours_GAD_cor,
      DE_hours_GAD_cor,
      UK_hours_GAD_cor) %>%

rename(r = estimate, t = statistic) %>% select(-method, -alternative) %>%
mutate(R2 = r^2) %>%
mutate(Country = c("USA", "DEU", "UK")) %>%
select(Country, r, R2, t, p.value, parameter, conf.low, conf.high)
```

GAD-7 and Hours Gaming - Correlation Summary

Country	r	`$r^2$`	t-statistic	p	df	lower 95% ci	upper 95% ci
USA	0.078	0.006	4.530	0.000	3380	0.044	0.111
DEU	0.119	0.014	3.778	0.000	1002	0.057	0.179
UK	0.073	0.005	2.004	0.045	741	0.002	0.145



Convergent Validity

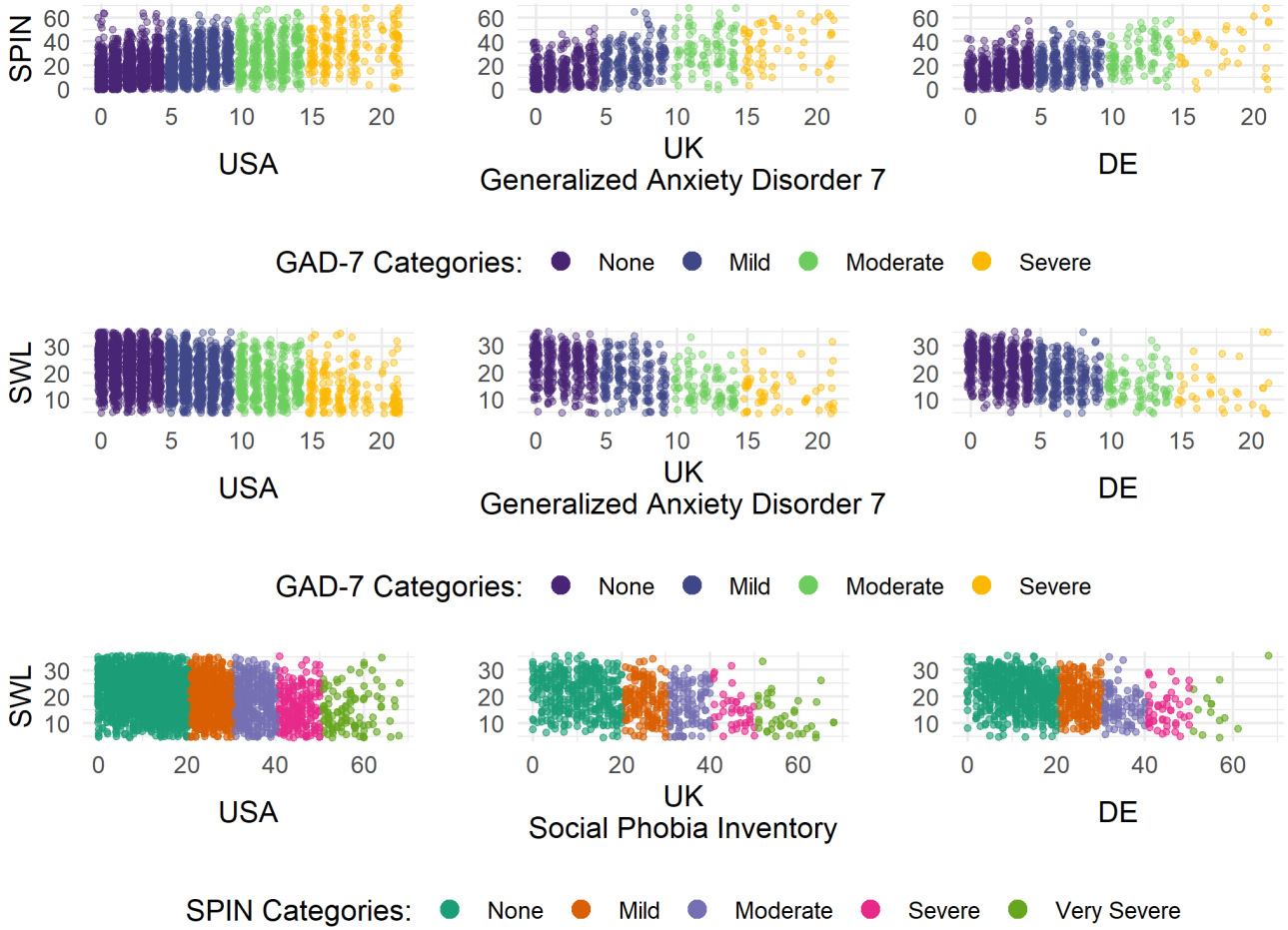
See R script R/03-additional-R-scripts/04-00-Correlation-by-country.R

Convergent Validity - Anxiety Measures

	Measure 1	Measure 2	r	p
Germany	Generalised Anxiety Disorder 7 Item Scale	Satisfaction With Life	-0.407	0
UK	Generalised Anxiety Disorder 7 Item Scale	Satisfaction With Life	-0.460	0
USA	Generalised Anxiety Disorder 7 Item Scale	Satisfaction With Life	-0.398	0
Germany	Generalised Anxiety Disorder 7 Item Scale	Social Phobia Inventory	0.464	0
UK	Generalised Anxiety Disorder 7 Item Scale	Social Phobia Inventory	0.526	0
USA	Generalised Anxiety Disorder 7 Item Scale	Social Phobia Inventory	0.463	0
Germany	Social Phobia Inventory	Satisfaction With Life	-0.347	0
UK	Social Phobia Inventory	Satisfaction With Life	-0.355	0
USA	Social Phobia Inventory	Satisfaction With Life	-0.305	0

Complex Data Visualisation

See R scripts R/03-additional-R-scripts/04-01 : 04-05



Open Science

What's the downside?

Downside(s) and Upside(s)

Open science requires additional effort on the part of the researcher.

1. Lots of data created (no longer an output file)
    - Several individual files; scripts, images, tables etc.
2. Significant effort and time
3. Discourages manual interaction with data

- Time + Experience = well-established workflow
- Reproducible = **Re-usable**
- Data sets and projects that can be examined thoroughly
    - Documentation of whole analysis process
- Discourages manual interaction with data