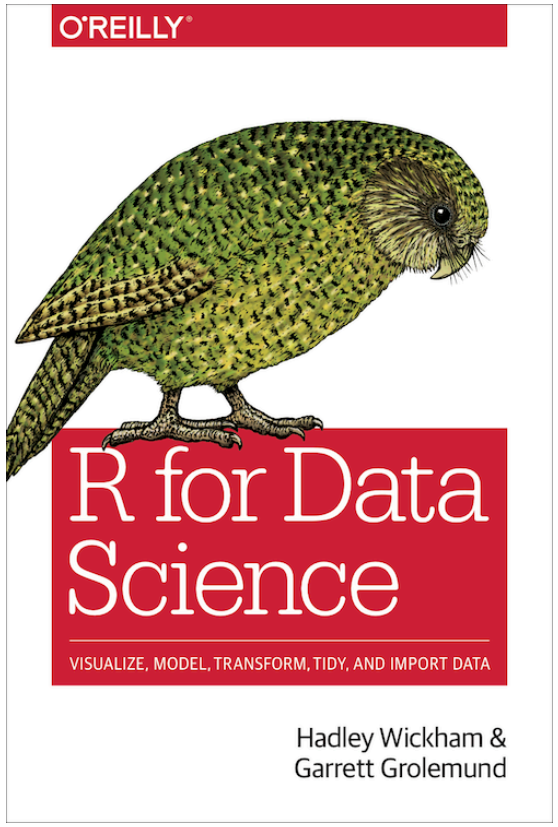
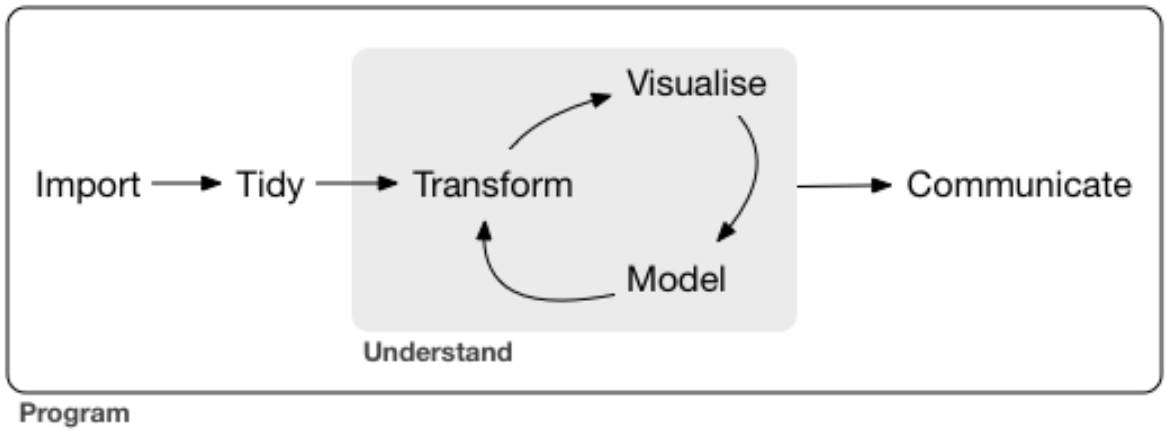
Yan Holtz | Brisbane | 2018

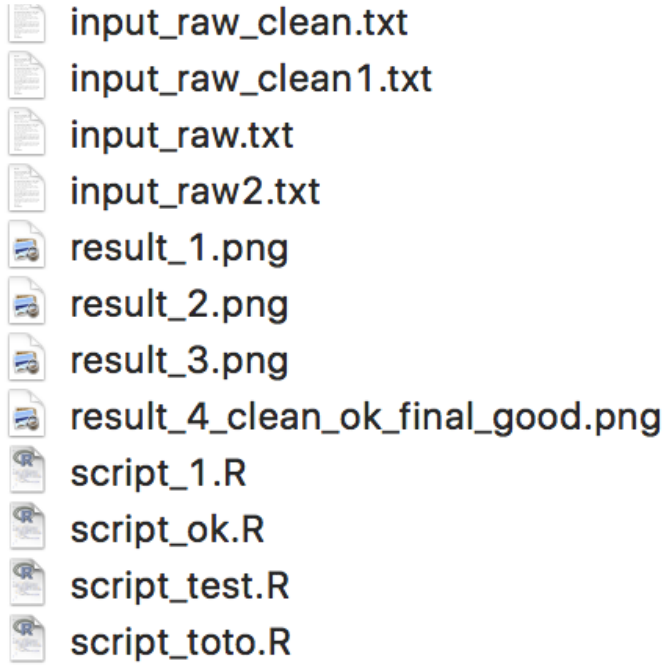
TEN TIPS FOR A BETTER DATA ANALYSIS PIPELINE IN RESEARCH





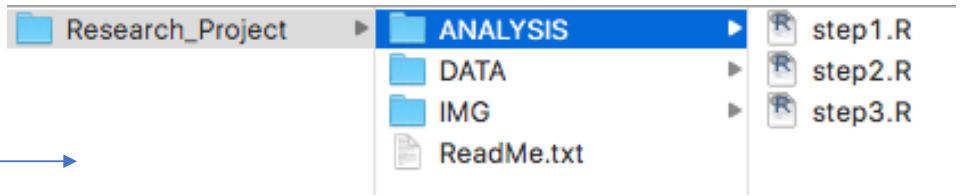
r4ds.had.co.nz

ORGANIZE


input_raw_clean.txt
input_raw_clean1.txt
input_raw.txt
input_raw2.txt
result_1.png
result_2.png
result_3.png
result_4_clean_ok_final_good.png
script_1.R
script_ok.R
script_test.R
script_toto.R

ORGANIZE

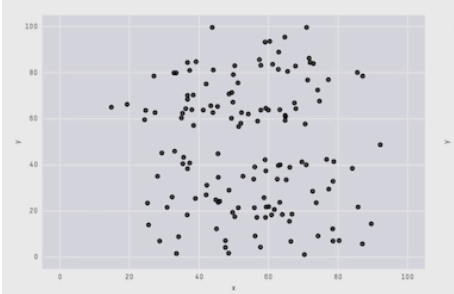
- input_raw_clean.txt
- input_raw_clean1.txt
- input_raw.txt
- input_raw2.txt
- result_1.png
- result_2.png
- result_3.png
- result_4_clean_ok_final_good.png
- script_1.R
- script_ok.R
- script_test.R
- script_toto.R



- Don’t modify input files manually
- Don’t copy and paste code / results
- Use code to do everything

CODE

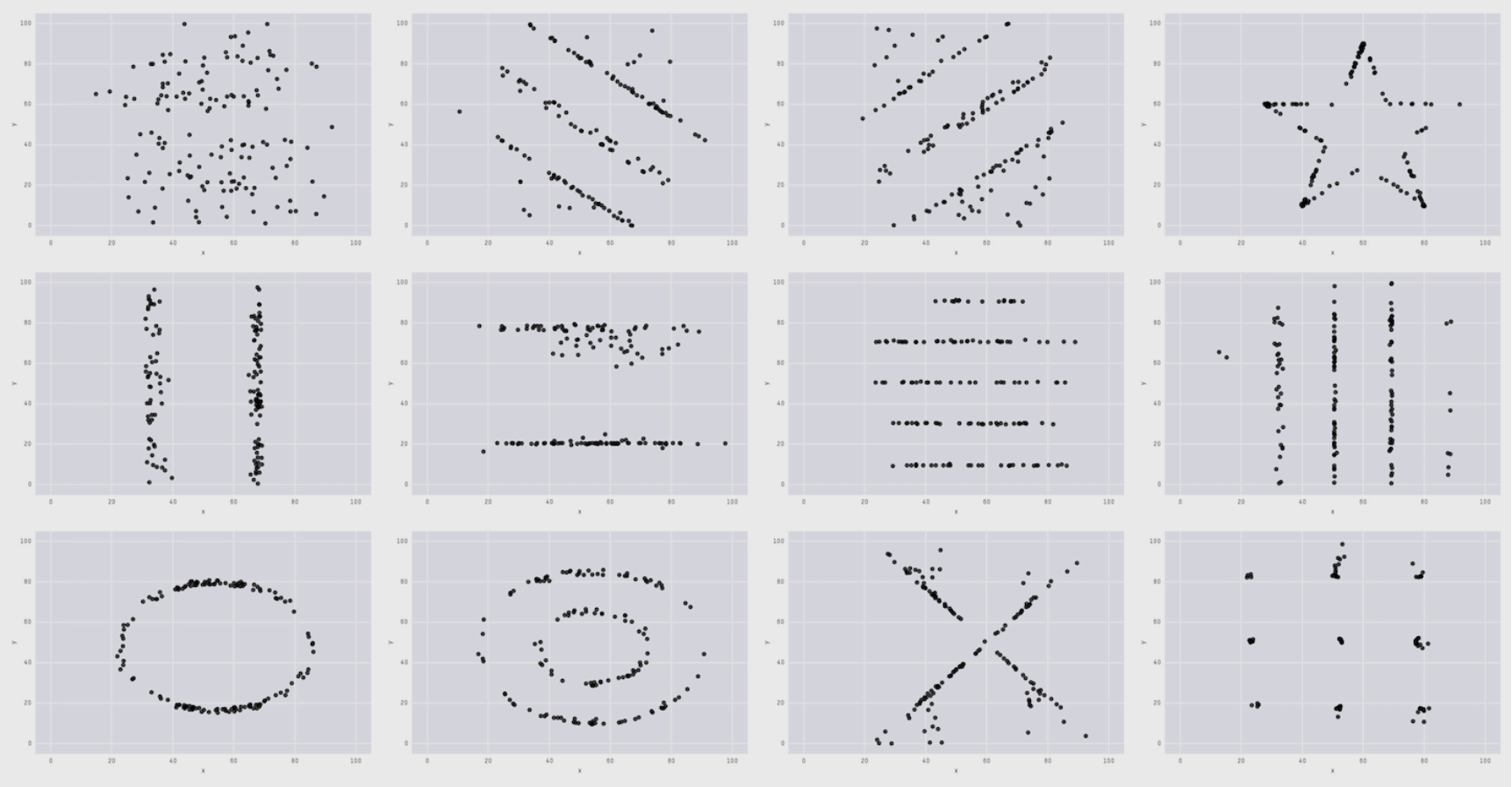
VISUALIZE



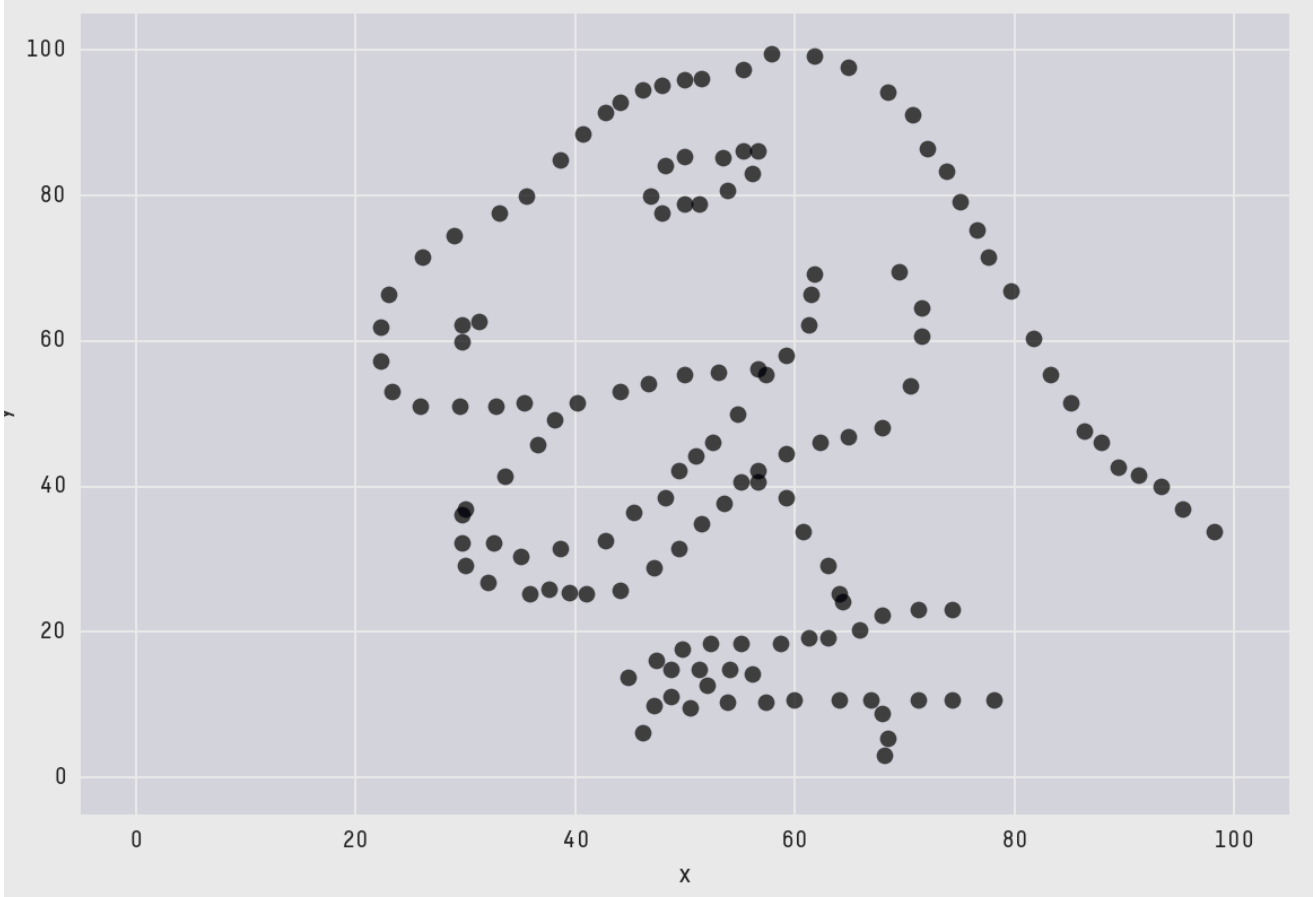
X Mean:	54.26
Y Mean:	47.83
X Sd:	15.76
Y Sd:	26.93
Correlation:	-0.06

VISUALIZE

X Mean:	54.26
Y Mean:	47.83
X Sd:	15.76
Y Sd:	26.93
Correlation:	-0.06



VISUALIZE



The Datasaurus Dozen

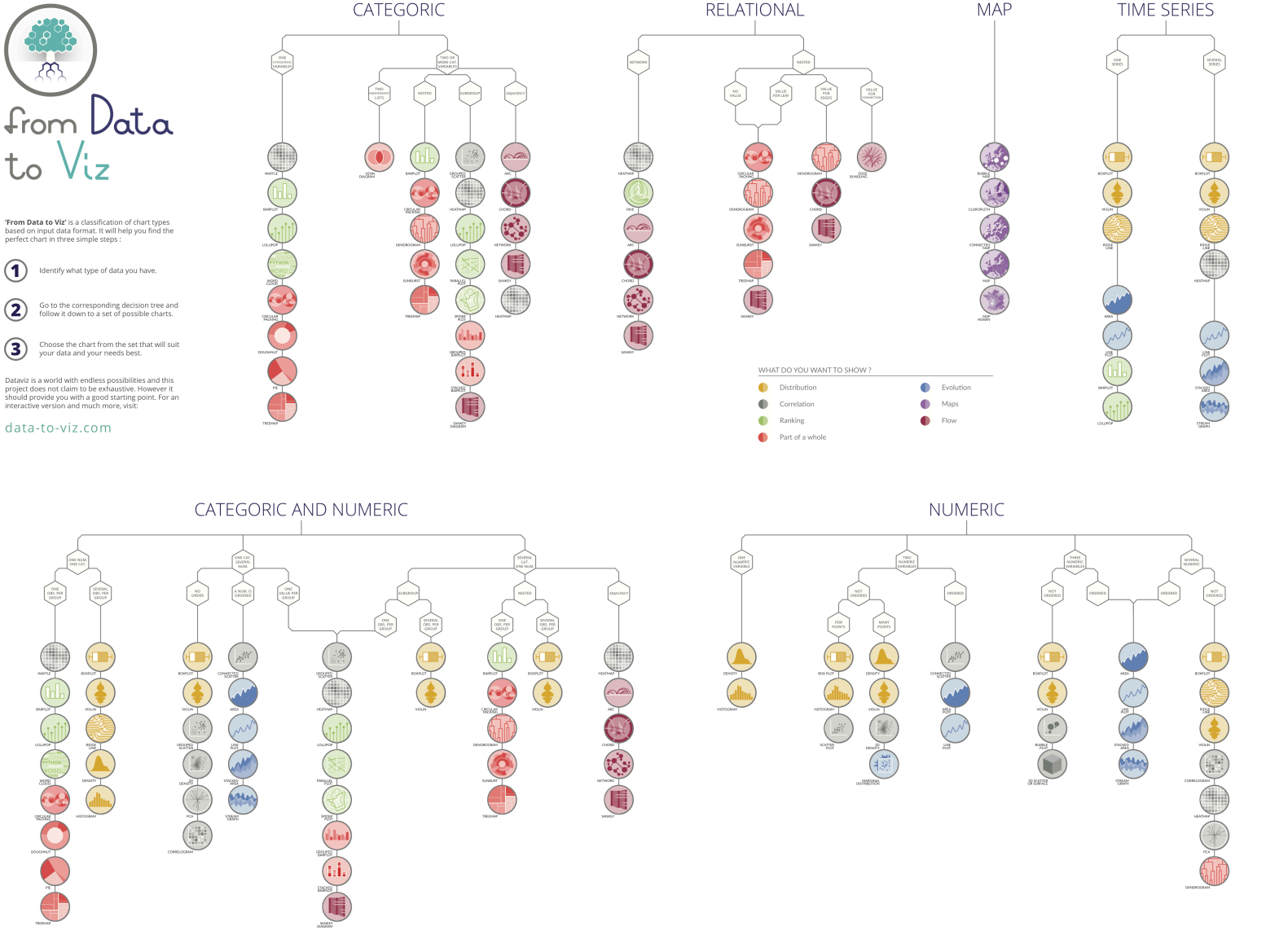
Cairo , 2017
Matejka & Fitzmaurice, 2017

AVOID CAVEATS

data-to-viz.com

R-graph-gallery.com

Python-graph-gallery.com



REPORT

R Markdown

from RStudio

```
53
54  # An interactive manhattan plot
55  ***
56
57  Using `HTML` outputs you can embed some interactive graphics. For example, the
    plotly library can transform any of your ggplot2 graphic in an interactive
    chart:
58
59  ```{r, message=FALSE, warning=FALSE, echo=FALSE}
60  # Libraries
61  library(plotly)
62  library(tidyverse)
63
64  # Prepare the dataset
65  don <- gwasResults %>%
66  ```
67
```

Title

Text

Code

Is your code reusable ?

REPORT

R Markdown

from RStudio

```
53
54  # An interactive manhattan plot
55  ***
56
57  Using `HTML` outputs you can embed some interactive graphics. For example, the
    plotly library can transform any of your ggplot2 graphic in an interactive
    chart:
58
59  ```{r, message=FALSE, warning=FALSE, echo=FALSE}
60  # Libraries
61  library(plotly)
62  library(tidyverse)
63
64  # Prepare the dataset
65  don <- gwasResults %>%
66  ```
67
```

Title

Text

Code

Program in Complex Trait Genomics

1 Loading data

2 Show data in a table

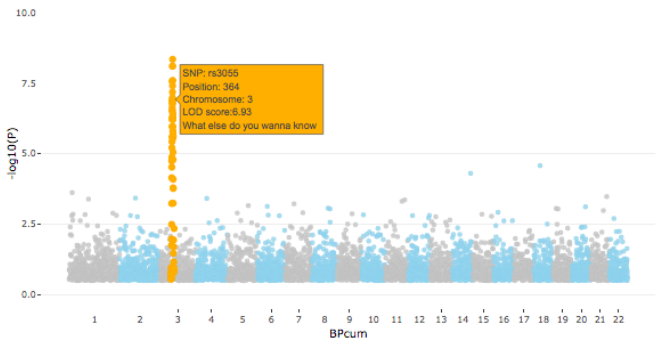
3 An interactive manhattan plot

3 An interactive manhattan plot

Using `HTML` outputs you can embed some interactive graphics. For example, the plotly library can transform any of your ggplot2 graphic in an interactive chart:

HIDE

```
# Make the plot
p <- ggplot(don, aes(x=BPcum, y=-log10(P), text=text)) +

    # Show all points
    geom_point( aes(color=as.factor(CHR)), alpha=0.8, size=1.3) +
    scale_color_manual(values = rep(c("grey", "skyblue"), 22 )) +

    # custom X axis:
    scale_x_continuous( label = axisdf$CHR, breaks= axisdf$center ) +
    scale_y_continuous(expand = c(0, 0) ) +     # remove space between plot area and x axis

    # Add highlighted points
    geom_point(data=subset(don, is_highlight=="yes"), color="orange", size=2) +

    # Custom the theme:
    theme_bw() +
    theme(
      legend.position="none",
      panel.border = element_blank(),
      panel.grid.major.x = element_blank(),
      panel.grid.minor.x = element_blank()
    ) +
    ylim(0, 10)

ggplotly(p, tooltip="text")
```



Interactive report

Is your code reusable ?

VERSION CONTROL

GitHub

Holtz Yan
holtzy

Data Analyst with a + in data visualization

Edit bio

Queensland Brain Institute
Brisbane, Australia
yan.holtz.data@gmail.com
https://holtzyan.wordpress.co...

Overview | Repositories 28 | Stars 22 | Followers 88 | Following 0

Pinned repositories

The-Python-Graph-Gallery
A website displaying hundreds of charts made with Python
Python ★122 15

R-graph-gallery
A website that displays hundre
R ★28 13

Pimp-my-rmd
A few tips about R markdown
HTML ★11 2

epuRate
A clean R Markdown template
HTML

GenMap-Comparator
An application to compare genetic maps with D3 & Shiny
R ★8 5

Publication-WSSMV-Re
Reproducible analysis of the p
resistance of Durum wheat to
HTML

452 contributions in the last year

May Jun Jul Aug Sep Oct Nov Dec Jan Feb
Mon
Wed
Fri

Learn how we count contributions.

holtzy / the-NB-COMO-Project

Unwatch 1 | Star 0 | Fork 0

Code | Issues 0 | Pull requests 0 | Projects 0 | Wiki | Insights | Settings

A Shiny app describing comorbidity in the Danish Health Register

Edit

Add topics

20 commits | 1 branch | 0 releases | 1 contributor

Branch: master | New pull request | Create new file | Upload files | Find file | Clone or download

holtzy Change 2 words — Latest commit e1a433e 21 days ago

DATA	Change 2 words	21 days ago
rsconnect/shinyapps.io/holtzyan	Change 2 words	21 days ago
www	John Correction	5 months ago
.DS_Store	John Correction	5 months ago
.Rhistory	First commit	5 months ago
README.md	Last version before sending to collaborators	5 months ago
global.R	Change 2 words	21 days ago
server.R	Change 2 words	21 days ago
ui.R	Change 2 words	21 days ago

README.md

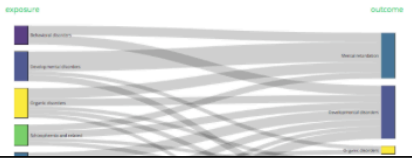
The NB-COMO Project

Overview

The NB-BOMO project aims to explore the patterns of comorbidity within treated mental disorders. It explores different ways to capture the complex patterns of comorbidity, notably through data visualization techniques. The first part of this project explores COMO within the Danish National Patient Registry (DNPR), one of the world's oldest nationwide hospital registries.

This repository gives the code of a web application that allows to interactively explore our results. It goes along with our peer reviewed publication (work in progress).

Here is a screenshot of one of the multiple visualizations proposed in the website:



SHARE

Acknowledgment

John McGrath

Naomi Wray
Peter Vissher
Jian Yang

Contact

@R_Graph_Gallery

github.com/holtzy/Talk

Yan.holtz.data@gmail.com

R-graph-gallery.com

Program in Complex Trait Genomics

NIELS BOHR PROFESSORSHIP PSYCHIATRIC EPIDEMIOLOGY

The University Of Queensland AUSTRALIA | IMB Institute for Molecular Bioscience

The University Of Queensland AUSTRALIA | QBI Queensland Brain Institute