

# Elevating Data Analysis



A roadmap for researchers

Yan Holtz | 🇦🇺, 🇫🇷 & 🌐 | March 2024

Hi, I'm Yan 🖐️

🎓 6 years in Academia

💻 5 years as a Software Engineer



# I ❤️ Dataviz & Code

→ [R-graph-gallery.com](https://r-graph-gallery.com)

→ [Data-to-viz.com](https://data-to-viz.com)

## Correlation



Scatterplot



Heatmap



Correlogram



Bubble



2D Density



Connected Scatter

## Distribution



Violin



Density



Histogram



Boxplot



Ridgeline

## Ranking



Barplot



Wordcloud



Parallel



Lollipop



Circular Barplot



Spider / Radar

## Part Of A Whole



Treemap



Donut



Pie Chart



Dendrogram



Voronoi



Circular Packing

## Evolution



Line chart



Area chart



Stacked Area



Streamgraph



Timeseries

## Map



Map



Choropleth



Hexbin



Cartogram



Connection



Bubble

## Flow



Chord Diagram



Network



Sankey



Arc Diagram



Edge Bundling

I remember the good old days!

**Manual formatting**

**Did you edit this file  
in Excel?**

Manual to

Did you edit this file in  
Excel?

Manual formatting

**Reading a script 1 week later**

Did you edit this file in  
Excel?

Reading a file 1 week

Manual formatting

**Copy chart to report again and again**



Did you edit this file in  
Excel?

**Send email. Nooooo it's wrong!**

**Where is this ~/Tom/Tmp/input.csv file again?**

Did you edit this file in  
Excel?

Manual formatting

Reading a file 1 week ago

Send email. Nooooo it's wrong!

Copy chart to ...

Where is this ~/Tom/Tmp/input.csv file again?

Did you edit this file in  
Excel?

Manual formatting

Reading a file 1 week later  
Nooooo it's wrong!

Send email

Copy chart

**Boss edits script. So do I.**

Where is this ~/Tom/Tmp/input.csv file again?

Did you edit this file in  
Excel?

Manual formatting

Reading a file 1 w  
Nooooo it's wrong!

Send er Boss

**Wait. script\_final or script\_final\_real.R?**

copy script. So do I.

Where is this ~/Tom/Tmp/input.csv file again?

Did you edit this file in  
Excel?

Manual formatting  
Reading a file 1 was  
10000 it's wrong!

**Kate internship's over. Anyone knows how to  
run her analysis?**

Wait. script\_tmp.  
copy



# How can we make it better?



Productivity



Visibility



Reproducibility



Collaboration

# 3h Roadmap 🤔 ?

① Better code



② Report from code



③ Store online



## 1 Better code → Ditch setwd

```
# Set the working directory  
setwd("/Users/holtzy/Desktop/")
```

```
# Ask R to read a file that is located in this directory  
data <- read.csv("data.csv")
```

## 1 Better code → Ditch setwd

```
# Set the working directory  
setwd("/UseX/holtzy/Desktop/")
```

```
# Ask R to read a file that is located in this directory  
data <- read.csv("data.csv")
```



## 1 Better code → Auto formatting

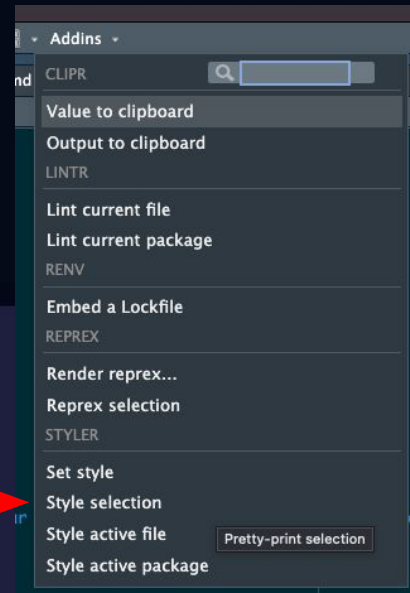
```
cleaNDatA = data %>%  
  rename(ID = 1, Name = 2, Age = 3) %>%  
  mutate(Age = ifelse(Age < 0, NA, Age)) %>% drop_na() %>%  
  → mutate(Age = as.integer(Age)) %>%  
  select(-Name)
```

The diagram illustrates the transformation of messy R code into clean, auto-formatted code. Red arrows indicate the mapping between the original code and the formatted code:

- Two arrows point from the original `cleaNDatA` and `data` to the formatted `cleaNDatA = data`.
- An arrow points from the original `rename(ID = 1, Name = 2, Age = 3)` to the formatted `rename(ID = 1, Name = 2, Age = 3)`.
- An arrow points from the original `mutate(Age = ifelse(Age < 0, NA, Age))` to the formatted `mutate(Age = ifelse(Age < 0, NA, Age))`.
- An arrow points from the original `drop_na()` to the formatted `drop_na()`.
- An arrow points from the original `mutate(Age = as.integer(Age))` to the formatted `mutate(Age = as.integer(Age))`.
- An arrow points from the original `select(-Name)` to the formatted `select(-Name)`.

# 1 Better code → Auto formatting

```
clean_data <- data %>%  
  rename(ID = 1, Name = 2, Age = 3) %>%  
  mutate(Age = ifelse(Age < 0, NA, Age)) %>%  
  drop_na() %>%  
  mutate(Age = as.integer(Age)) %>%  
  select(-Name)
```



## 1 Better code → Functions

```
print(
  round(mean(subset(
    na.omit(data), species == "Adelie" & island == "Torgersen"
  )$bill_length_mm), 2)
)
print(
  round(mean(subset(
    na.omit(data), species == "Adelie" & island == "Biscoe"
  )$bill_length_mm), 2)
)
print(
  round(mean(subset(
    na.omit(data), species == "Adelie" & island == "Dream"
  )$bill_length_mm), 2)
)
```

## 1 Better code → Functions

```
calc_mean_bill <- function(island_name) {  
  filtered_data <- subset(na.omit(data), species == "Adelie" & island == i  
  mean_bill_length <- mean(filtered_data$bill_length_mm)  
  return(round(mean_bill_length, 2))  
}
```

```
# Call the function for each island  
calc_mean_bill("Torgersen")  
calc_mean_bill("Biscoe")  
calc_mean_bill("Dream")
```



{styler}



1 Better code

shortcut



comments

function()

2 Report from code with



Demo



# 😞 4 problems remaining

🔥 Safety


✉️ Sharing

👤 Collaboration

🕒 History




# GitHub is a code hosting platform.

[Product](#) [Solutions](#) [Open Source](#) [Pricing](#)

List of ALL projects

Search or jump to... 7 Sign in Sign up





**Holtz Yan**  
holtzy  
Follow


I ❤️ data visualization 🧑  
3.6k followers · 8 following  
Software Engineer - Dataviz  
Montpellier, France  
<https://www.yan-holtz.com>  
[@R\\_Graph\\_Gallery](#)  
<https://datavizuniverse.substack.com>  
Achievements


Overview Repositories 62 Projects Packages Stars 80


Pinned


 **data\_to\_viz** Public  
Leading to the dataviz you need  
HTML 894 278

 **D3-graph-gallery** Public  
A collection of simple graphics made with D3.js  
HTML 766 238

 **The-Python-Graph-Gallery** Public  
A website displaying hundreds of charts made with Python  
HTML 1.7k 353

 **R-graph-gallery** Public  
A website that displays hundreds of R charts with their code  
HTML 613 214

 **dataviz-inspiration** Public  
A collection of viz I love  
TypeScript 76 8

 **react-graph-gallery** Public  
A set of graph examples showing how to make react and d3.js work together  
TypeScript 259 19

History of ALL my code updates

1,549 contributions in the last year

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

Learn how we count contributions

Less More

2024  
2023  
2022  
2021  
2020

github handle → holtzy

profile description

Pinned repos



Demo



Subscriptions open

# Turn messy R scripts into clean reports

An **interactive mini course** helping R developers be so. much. more. efficient at data analysis.



Taught by Yan Holtz





Subscriptions open

# Turn messy R code into clean reports

An interactive workflow for developers to be  
so much more productive in analysis.

Holtz

productive-r-workflow.com





data-to-viz.com

dataviz-inspiration.com

R-graph-gallery.com

Python-graph-gallery.com

D3-graph-gallery.com

React-graph-gallery.com

Yan-Holtz.com

@r\_graph\_gallery

**What can I do for you? 🤔**