



15 September 2022

# THE JOYS OF AN UNCONVENTIONAL PORTFOLIO

Amit Grinson

# Developing a portfolio



“Take a dataset from Kaggle and analyze it”



“Take a dataset from Kaggle and Build a dashboard”



“Tutorial Hell”



# But what Work...?



# About me

- Risk Data Analyst @ Payoneer
- Father & Partner
- Proud dog owner
- Enjoy data analysis & visualizations
  - Blog occasionally at [amitlevinson.com](http://amitlevinson.com)
- MA in Sociology & Anthropology

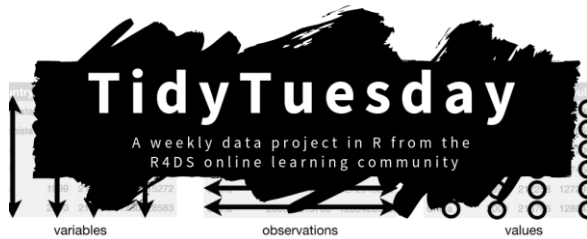


[www.milo-the-dog.amitlevinson.com](http://www.milo-the-dog.amitlevinson.com)



# Data visualization challenges

## TidyTuesday



<https://github.com/rfordatascience/tidytuesday>

## MapChallenge



[www.30daymapchallenge.com](http://www.30daymapchallenge.com)

## ChartChallenge



[www.30daychartchallenge.org](http://www.30daychartchallenge.org)

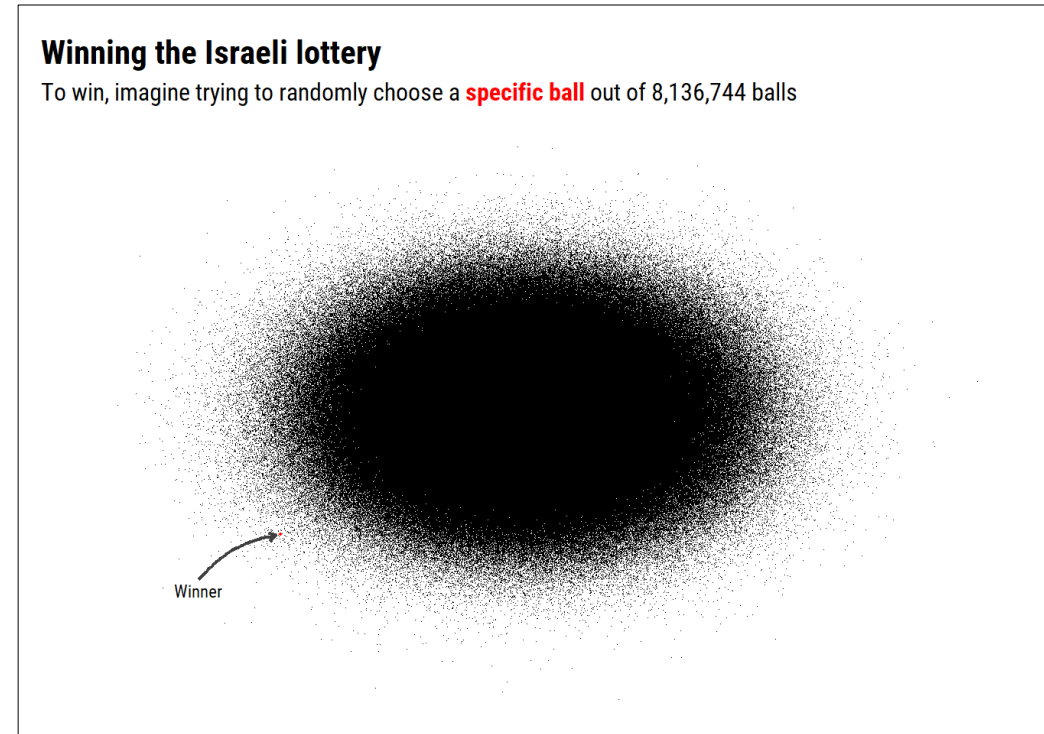
# 1. Working with various data



# Methods I want to learn

# "Defining uncertainty in the Israeli lottery"

- What is a binomial distribution?
- What are combinations?
- What is a geometric distribution?
- Visualizing winning
- Visualizing Money spent





# Methods I want to learn

“Defining uncertainty in the Israeli lottery”

- What is a binomial distribution?
- What are combinations?
- What is a geometric distribution?
- Visualizing winning
- Visualizing Money spent

Lottery probabilities with the geometric distribution		
Lottery probabilities winning by a given attempt, the money spent reaching there and your chances of winning by then		
Attempt	Money spent <sup>1</sup>	% winning by then
1	\$2	0.00001%
10	\$17	0.00012%
100	\$166	0.00123%
500	\$829	0.00614%
1,000	\$1,657	0.01229%
2,500	\$4,143	0.03072%
5,000	\$8,286	0.06143%
10,000	\$16,571	0.12282%
25,000	\$41,429	0.30678%
100,000	\$165,714	1.22147%
250,000	\$414,286	3.02576%
500,000	\$828,571	5.95997%
1,000,000	\$1,657,143	11.56473%
10,000,000	\$16,571,429	70.74129%
50,000,000	\$82,857,143	99.78557%
<sup>1</sup> Money spent (rounded up) corresponds to the cumulative number of attempts played		



# Methods I want to learn

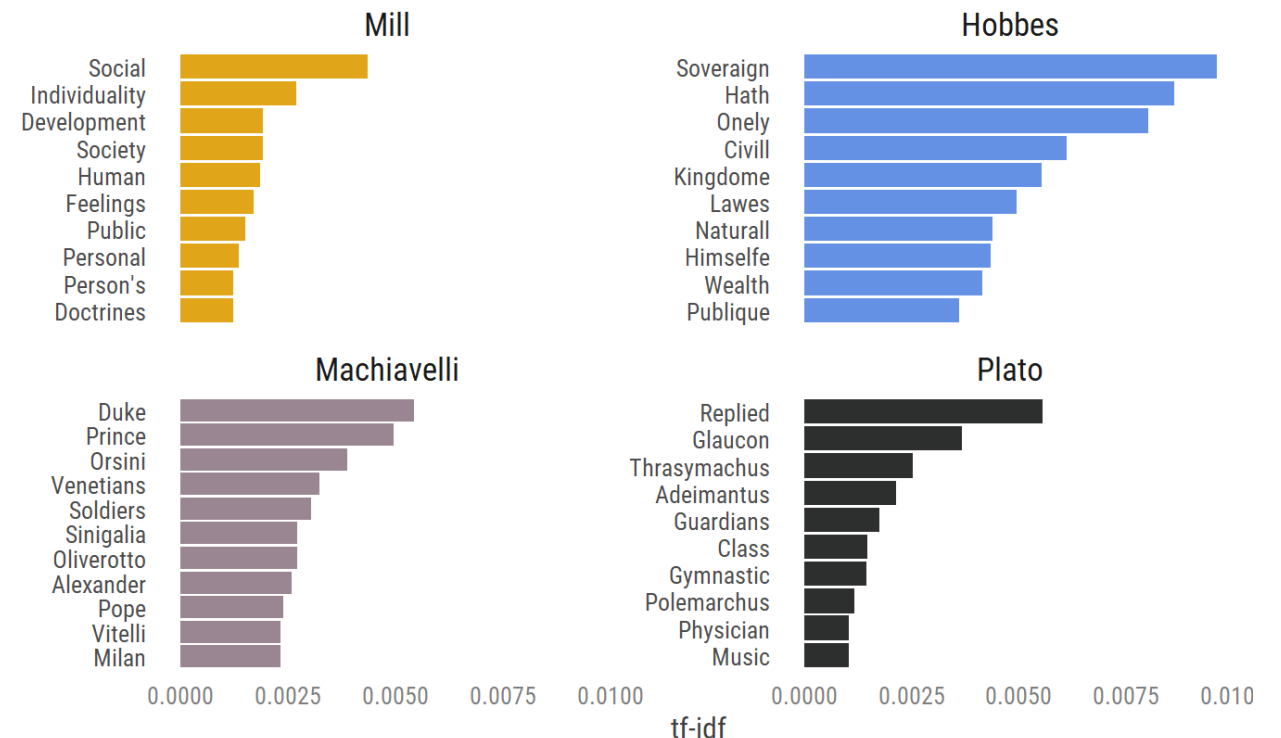
## "Learning TF-IDF with Political Theorists"

$$tf = \frac{\text{Number of times a word appears in a document}}{\text{Total number of words in that document}}$$

$$idf = \log \frac{N \text{ documents in corpus}}{n \text{ documents containing the term}}$$

### Term Frequency Inverse Document Frequency - Political theorists

tf-idf for The Leviathan (Hobbes), On Liberty (Mill), The Prince (Machiavelli) and Republic (Plato)



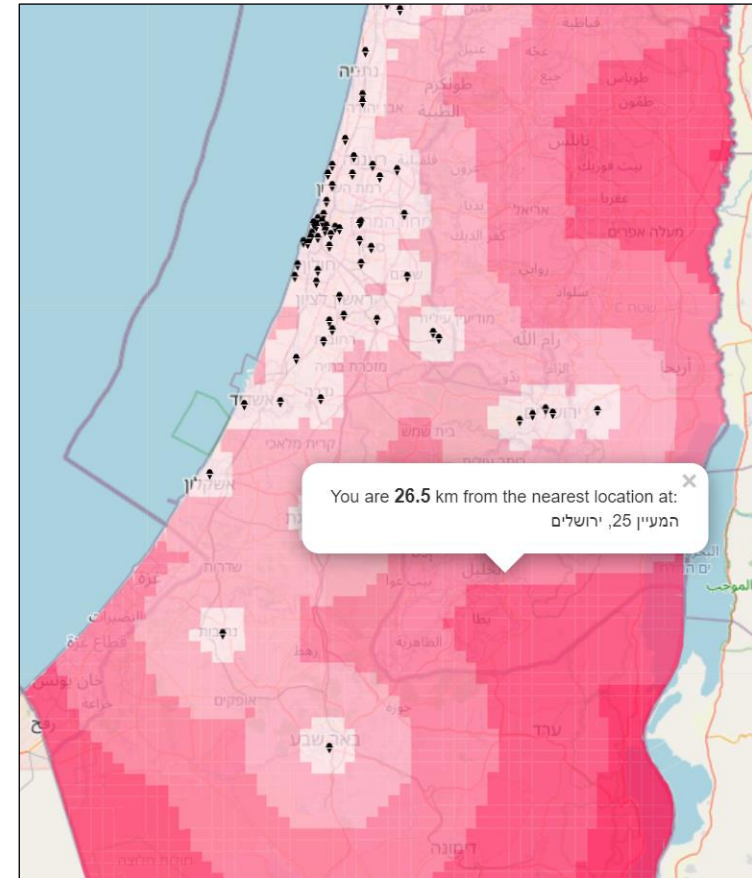
## 2. Taking complex ideas and breaking them down



# Messy data

## "Distances to Golda Ice-Cream Location"

- Scraping
- Cleaning
- Geocoding
- Reverse geocoding
- Checking discrepancies
- ...





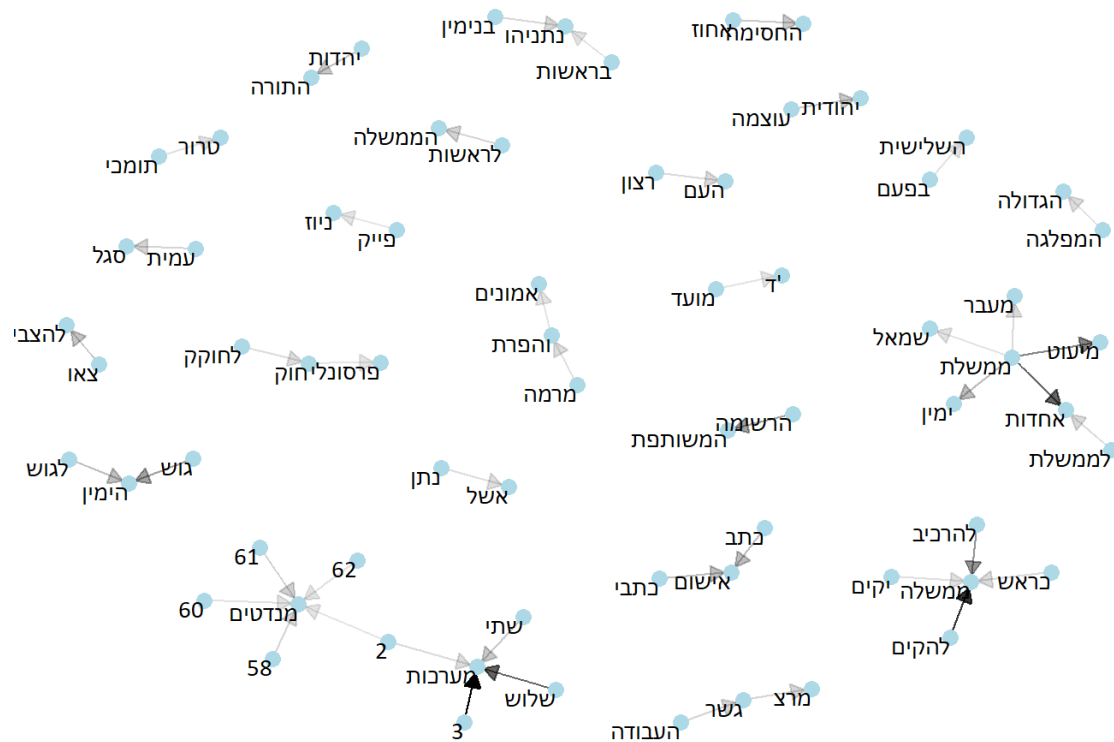




# Messy data

“Israeli elections on Twitter”

Twitter text bigram



“Happy families are all alike; every unhappy family is unhappy in its own way.” — Leo Tolstoy

**“Tidy datasets are all alike, but every messy dataset is messy in its own way.” —Hadley Wickham**





# 3. Working with messy data



# GitHub Actions

## "Automatic Rendering of a Plot with GitHub Actions"

How can I run something independent of my computer?

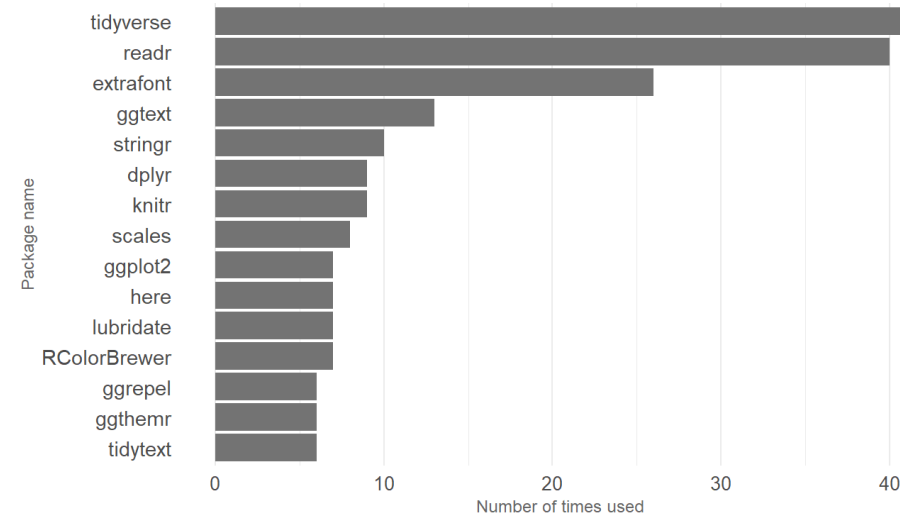
```
name: Render library update

on: [push]

jobs:
  build:
    runs-on: macOS-latest
    if: "contains(github.event.head_commit.message, 'initial commit')"
    steps:
      - uses: actions/checkout@v2
      - uses: r-lib/actions/setup-r@v1
      - name: Install packages
        run:
          Rscript -e "install.packages(c('tidyverse', 'showtext'))"
      - name: Render r plot file
        run:
          Rscript -e "source('extra/packages-plot.R')"
      - name: Commit results
        run: |
          git config --local user.email "actions@github.com"
          git config --local user.name "GitHub Actions"
          git add extra/packages-used.png
          git commit -m 'Re-build package plot' || echo "No changes to commit"
          git push origin || echo "No changes to commit"
```

### Frequently used packages in #TidyTuesday

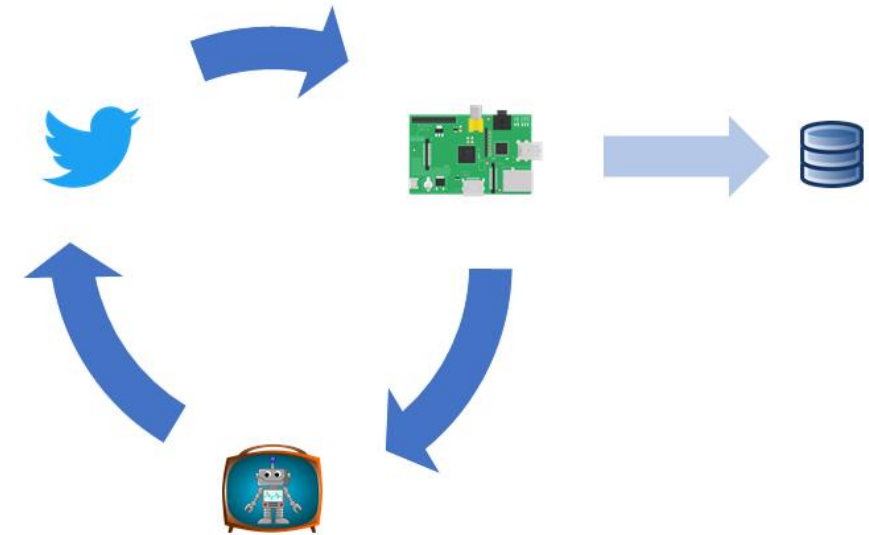
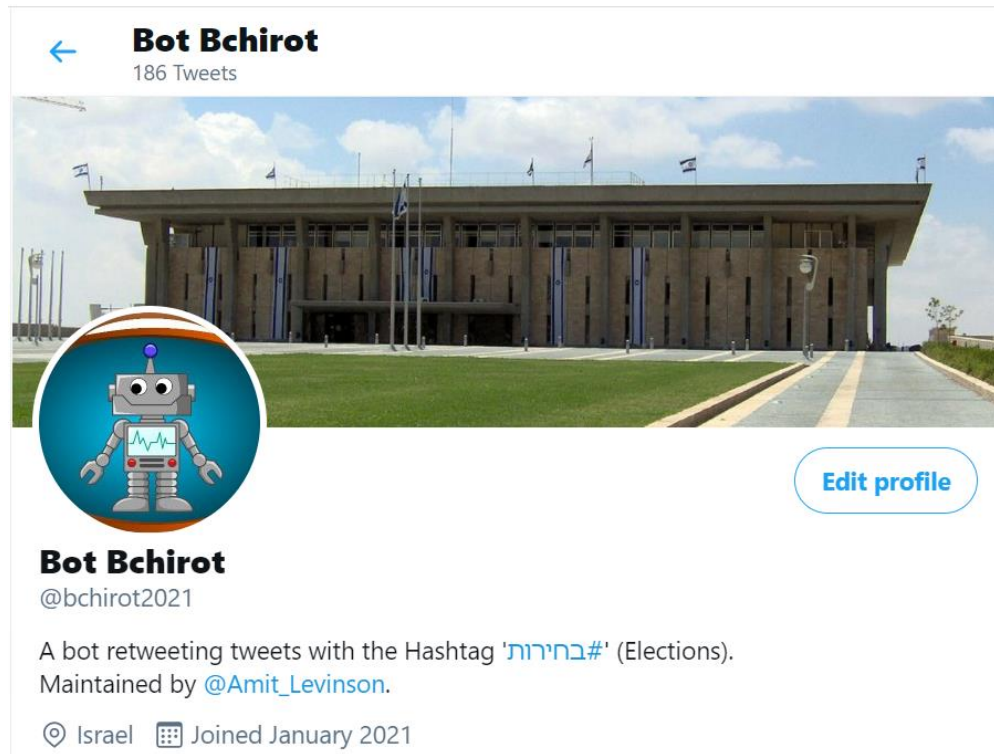
Plot is rendered on every 'initial commit' to this repository, showing my 15 most frequently used packages in #TidyTuesday



Total scripts: 53  
Last updated: Sep 04, 2022

# Raspberry Pi


## "Israeli Elections Twitter-bot"




# 4. Expanding my horizons








# Building a website

[BLOG](#) [TALKS](#) [PROJECTS](#) [VISUALIZATIONS](#) 



**Amit Grinson**  
Risk Data Analyst  
Payoneer

Hi! I'm Amit Levinson; I love to explore, analyze and visualize data. I currently live with my wife, our baby boy and our dog [Milo](#) in Tel-Aviv, Israel. My interests include Privacy & Trust, Data analytics, Visualizations, Statistics and [R](#). That's also what you'll probably find me writing about on this website 😊

Make sure to visit [R-bloggers](#) & [R-weely](#) to explore many more tutorials and resources. Enjoy!

### Interests

- Data analytics
- Storytelling and data visualization
- Trust & privacy

### Education

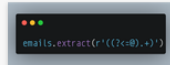
🎓 MA in Sociology & Anthropology, 2021  
Ben-Gurion University

🎓 BA in Sociology, Anthropology & Political Science, 2018  
Ben-Gurion University

## Blog Posts

### Text Manipulation & RegEx in a Data Analysis Flow (with examples in SQL vs Python/R)

Going over various cases of text operations in a data analysis



# 5. 'Selling' yourself



Working with  
various data

Taking complex  
ideas and  
breaking them  
apart

Working with  
messy data

Expanding my  
horizons

Selling yourself



# Choosing a project

- What am I **currently curious** about? What do I want to **learn**?
- What **challenge** did I recently **solve**?
- Do it **for myself, not for others**
- **Have fun!**





# Thank you!

*Many thanks to our generous sponsors for making this day a huge success!*

