

Turning Polars DataFrames into Pretty Pictures and Great Tables

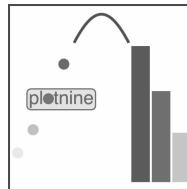
Jeroen Janssens

New York Open Statistical Programming Meetup - May 7, 2024



hvPlot

bok^okeh



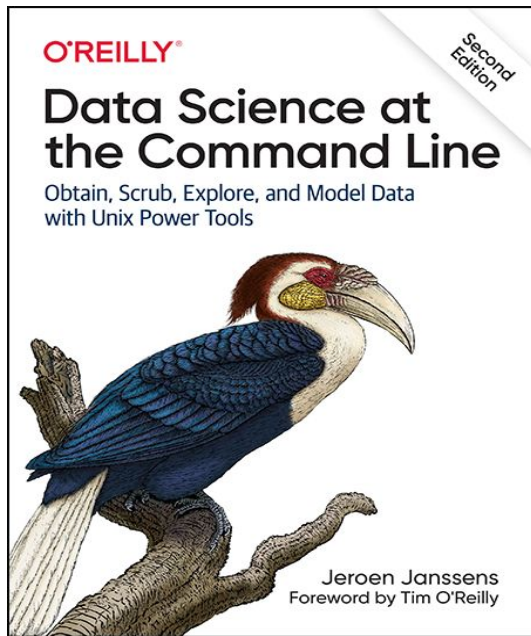
Great Tables

<https://github.com/jeroenjanssens/turning-polars-dataframes-into-pretty-pictures-and-great-tables>

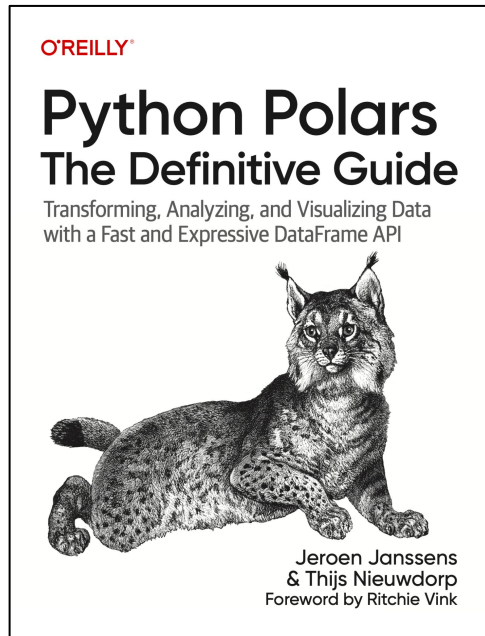
Jeroen Summarized in Three Books



2013



2014 & 2021



2024

Things I Love





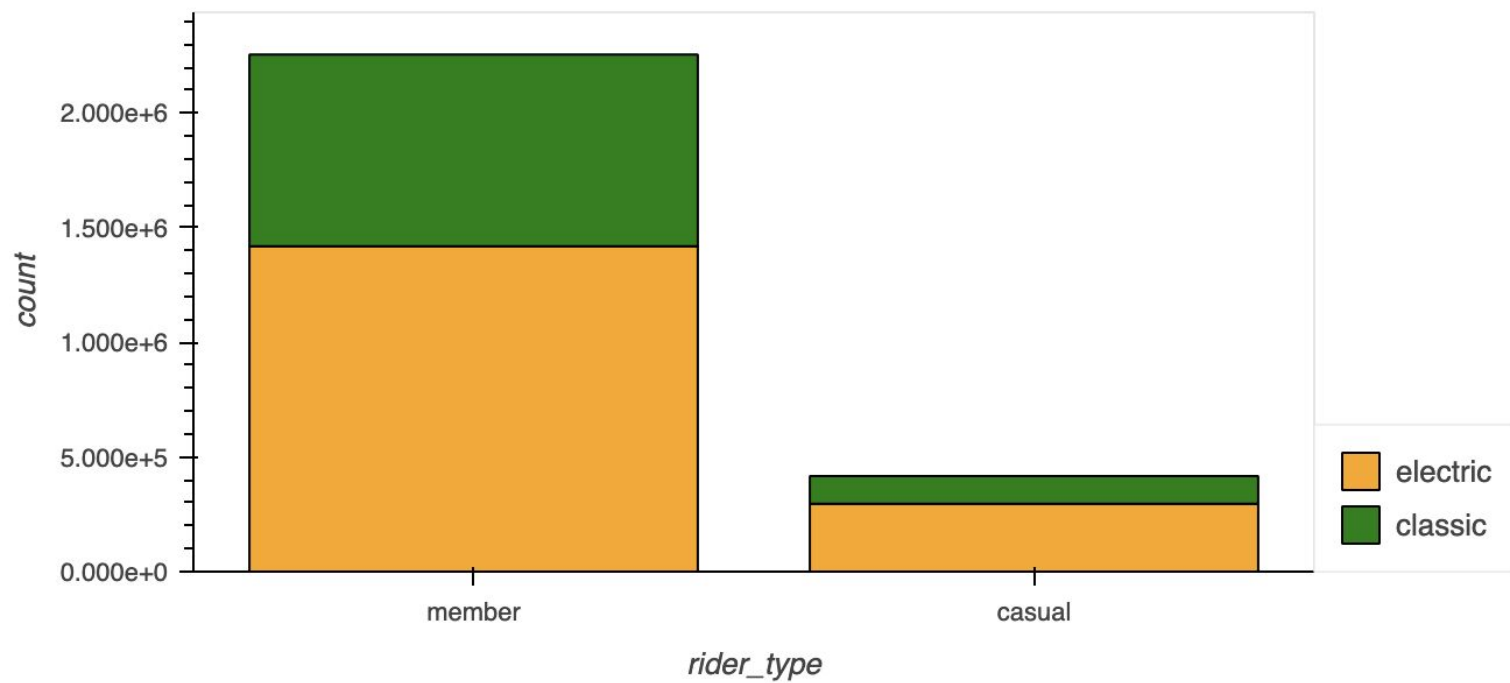
Our Goal

NYC Bike Trips March 2024

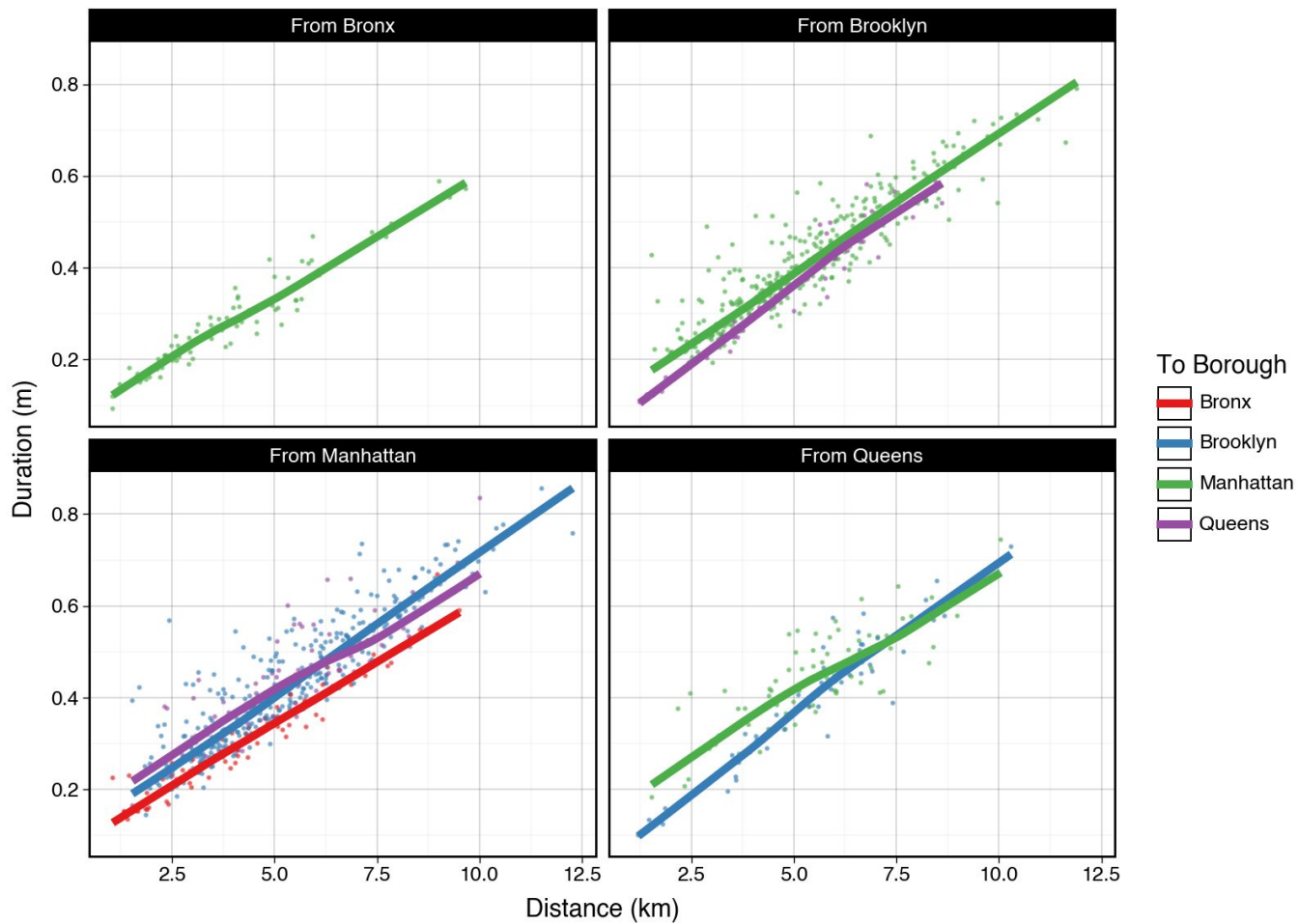
```
[1]: import polars as pl
      pl.read_parquet("data/biketrips/*.parquet")
```

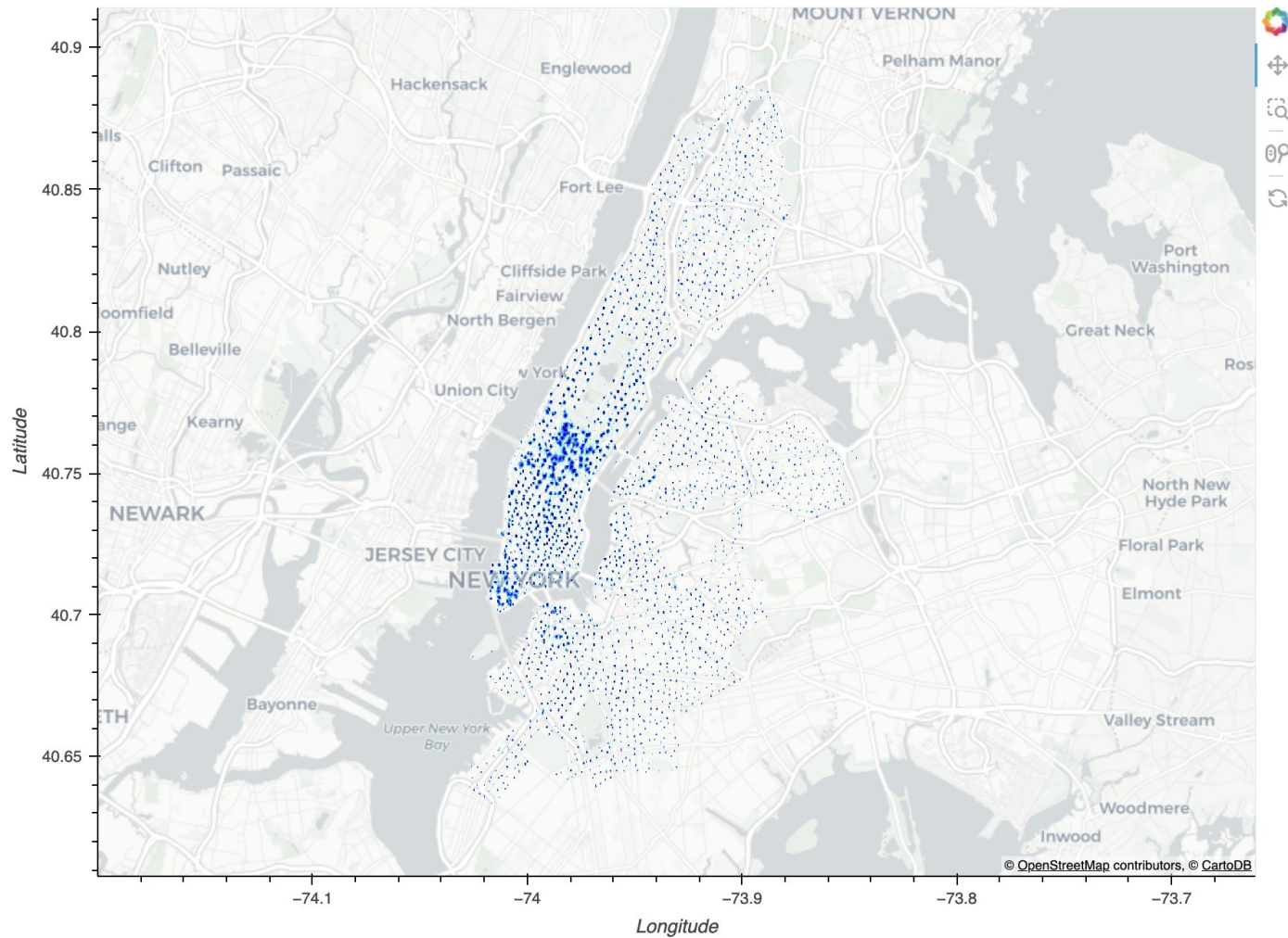
[1]: shape: (2_735_398, 16)

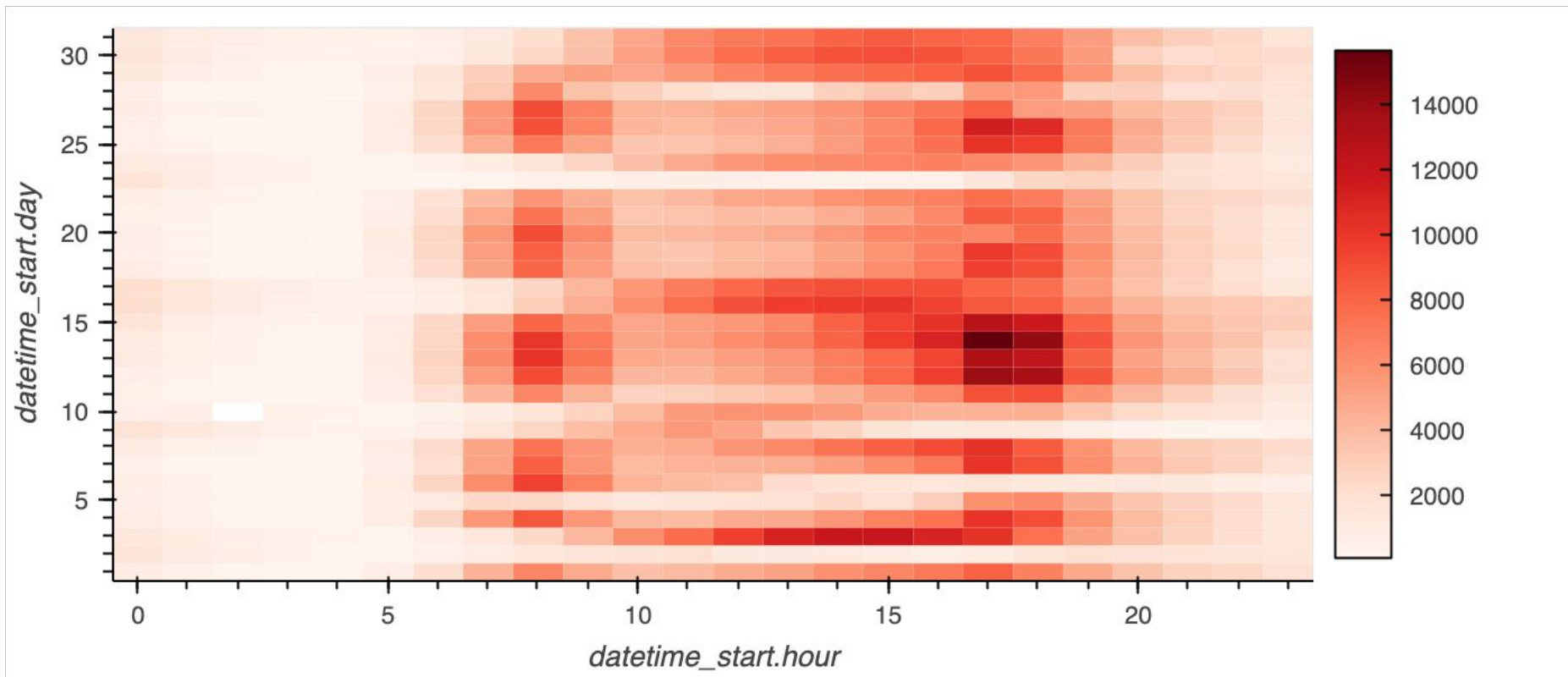
bike_type	rider_type	datetime_start	datetime_end	duration	station_start	station_end	neighborhood_start	neighborhood_end	borough_start	borough_end	lat_start	lon_start	lat_end	lon_end	distance
cat	cat	datetime[μs]	datetime[μs]	duration[μs]	str	str	str	str	str	str	f64	f64	f64	f64	f64
"electric"	"member"	2024-03-01 00:00:02	2024-03-01 00:27:39	27m 37s	"W 30 St & 8 Av...	"Maiden Ln & Pe...	"Chelsea"	"Financial Dist...	"Manhattan"	"Manhattan"	40.749614	-73.995071	40.707065	-74.007319	4.83703
"electric"	"member"	2024-03-01 00:00:04	2024-03-01 00:09:29	9m 25s	"Longwood Ave &...	"Lincoln Ave & ...	"Longwood"	"Mott Haven"	"Bronx"	"Bronx"	40.816459	-73.896576	40.810893	-73.927311	2.665806
"classic"	"member"	2024-03-01 00:00:05	2024-03-01 00:31:02	30m 57s	"W 30 St & 8 Av...	"Broadway & Mor...	"Chelsea"	"Financial Dist...	"Manhattan"	"Manhattan"	40.749653	-73.995208	40.705945	-74.013219	5.086562
"electric"	"casual"	2024-03-01 00:00:05	2024-03-01 00:03:34	3m 29s	"Broadway & W 2...	"Broadway & W 2...	"Midtown"	"Flatiron Distr...	"Manhattan"	"Manhattan"	40.746424	-73.988559	40.742869	-73.989186	0.398297
"electric"	"member"	2024-03-01 00:00:09	2024-03-01 00:11:42	11m 33s	"DeKalb Ave & F...	"6 St & 7 Ave"	"Bedford-Stuyve...	"Park Slope"	"Brooklyn"	"Brooklyn"	40.69067	-73.957559	40.668663	-73.979881	3.087634
...
"classic"	"member"	2024-03-31 23:59:54	2024-04-01 00:32:09	32m 15s	"Pier 40 - Huds...	"Amsterdam Ave ...	"SoHo"	"Upper West Sid...	"Manhattan"	"Manhattan"	40.727714	-74.011296	40.779668	-73.98093	6.313617
"electric"	"member"	2024-03-31 23:59:56	2024-04-01 00:02:20	2m 24s	"Rivington St &...	"Ave D & E 3 St...	"Lower East Sid...	"East Village"	"Manhattan"	"Manhattan"	40.718701	-73.983141	40.720828	-73.977932	0.49954
"electric"	"casual"	2024-03-31 23:59:57	2024-04-01 00:09:13	9m 16s	"12 Ave & W 40 ...	"Riverside Blvd...	"Hell's Kitchen...	"Upper West Sid...	"Manhattan"	"Manhattan"	40.760944	-74.002864	40.777507	-73.988886	2.185411
"electric"	"casual"	2024-03-31 23:59:57	2024-04-01 00:15:39	15m 42s	"Hart St & Wyc...	"Monroe St & Be...	"Bushwick"	"Bedford-Stuyve...	"Brooklyn"	"Brooklyn"	40.704865	-73.919904	40.685144	-73.953809	3.606664
"classic"	"casual"	2024-03-31 23:59:58	2024-04-01 00:01:38	1m 40s	"5 Ave & E 30 S...	"5 Ave & E 30 S...	"Midtown"	"Midtown"	"Manhattan"	"Manhattan"	40.745985	-73.986295	40.745985	-73.986295	0.0

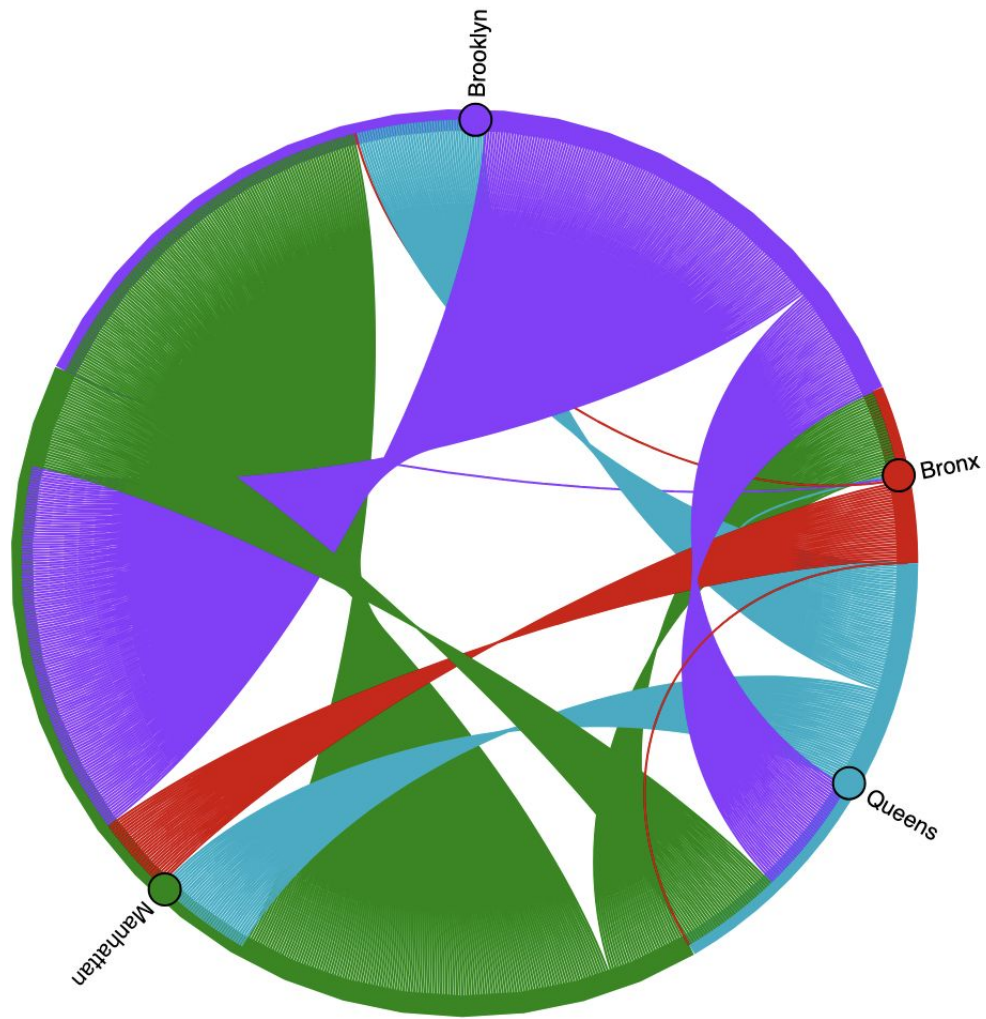


Trip distance and duration cross borough









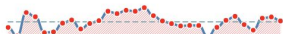









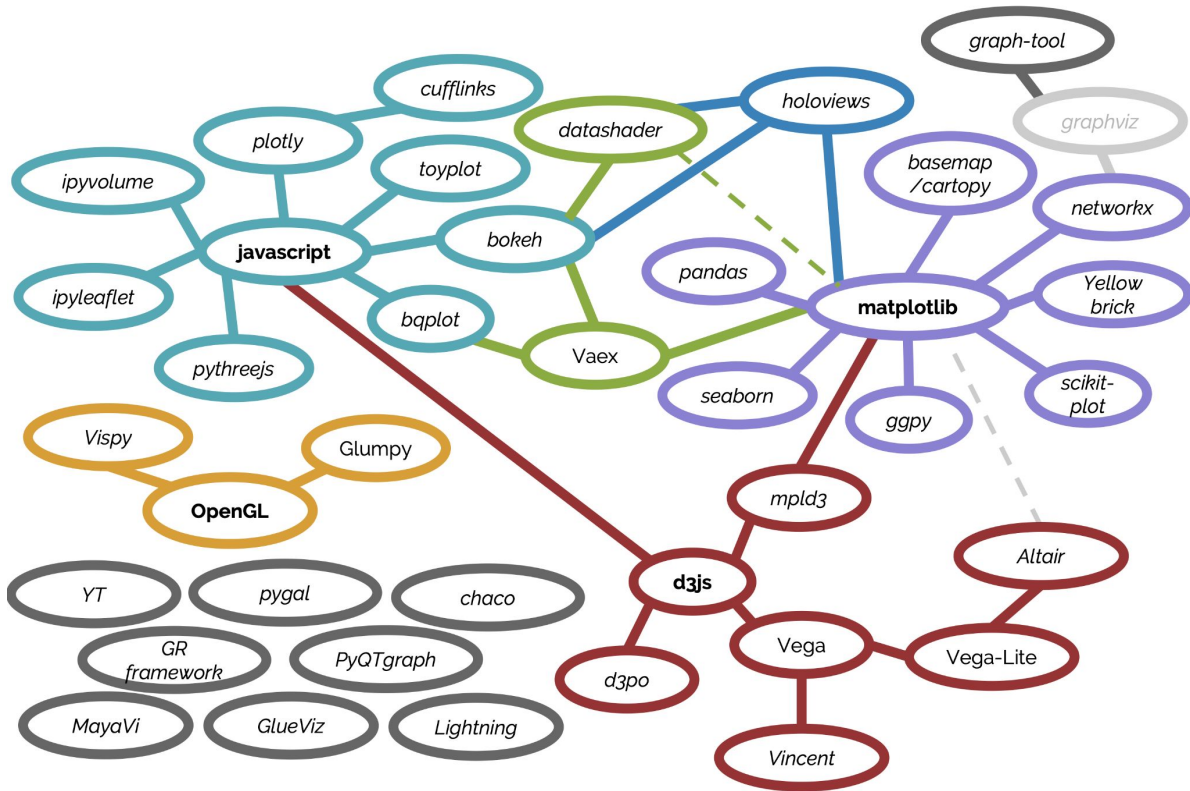


Busiest Bike Stations in NYC

In March 2024, Per Borough

Station	Neighborhood	Mean Daily Rides	Members	E-Bikes	Rides Per Day
Manhattan					
W 21 St & 6 Ave	Chelsea	354.2	91%	58%	
Forsyth St & Broome St	Lower East Side	322.2	95%	25%	
Broadway & W 58 St	Midtown	315.2	80%	70%	
Brooklyn					
Metropolitan Ave & Bedford Ave	Williamsburg	185.4	85%	68%	
N 7 St & Driggs Ave	Williamsburg	146.0	86%	66%	
Hanson Pl & Ashland Pl	Fort Greene	144.1	84%	61%	
Queens					
Queens Plaza North & Crescent St	Long Island City	127.0	85%	57%	
Vernon Blvd & 50 Ave	Long Island City	95.7	89%	66%	
31 St & Newtown Ave	Astoria	78.2	88%	60%	
Bronx					
Melrose Ave & E 150 St	Melrose	41.7	83%	89%	
E 161 St & River Ave	Concourse	35.6	75%	88%	
Plaza Dr & W 170 St	Mount Eden	31.7	84%	95%	
Source: NYC Citi Bike					

Data Visualization Landscape



Notebook Time

Notebooks

1. Introducing Polars expressions
2. Preparing the data
3. Visualizing the data

Notebooks are available at:

<https://github.com/jeroenjanssens/turning-polars-dataframes-into-pretty-pictures-and-great-tables>

Summary

1. Visualization landscape is like the Wild West
2. Built-in plot methods are fine
3. Many alternatives; Polars is not fully supported
4. Can always fall back to Pandas if needed
5. Tables matter too

Thank you

jeroenjanssens.com

jeroen@jeroenjanssens.com

[@jeroenhjanssens](#)