

Alexander Lex
<http://vdl.sci.utah.edu>



A Hodgepodge of Visualization Research: Provenance, User Studies, Misinformation



visualization
design lab





visualization
design lab

<http://vdl.sci.utah.edu/>



Visualization =

Human Data Interaction

visualization

**The purpose of computing is insight,
not numbers.**

pictures

[Card, Mackinlay, Shneiderman]

[Richard Wesley Hamming]

Banana *M. acuminata*

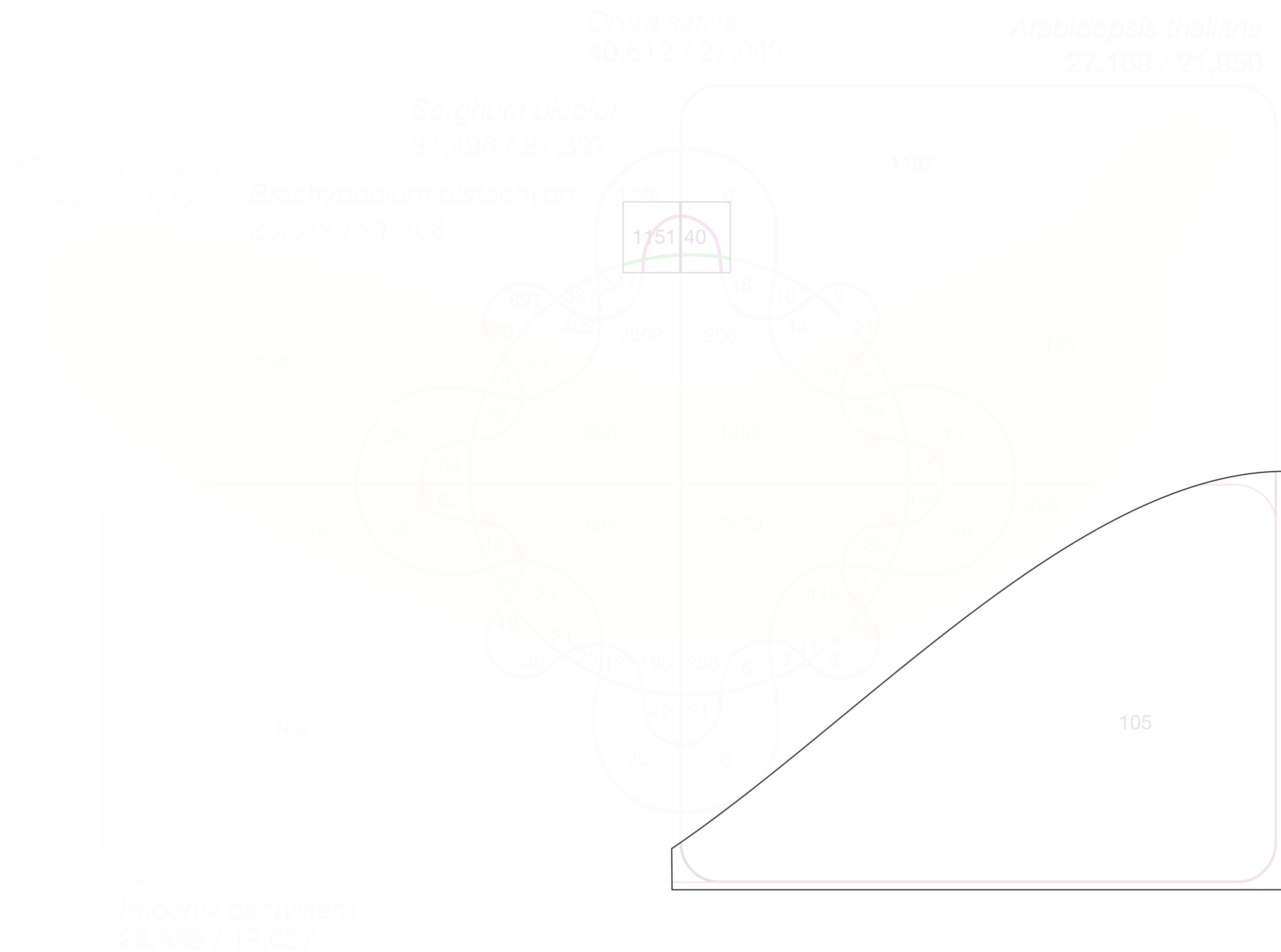
Date *P. dactylifera*

Cress *Arabidopsis thaliana*

Rice *Oryza sativa*

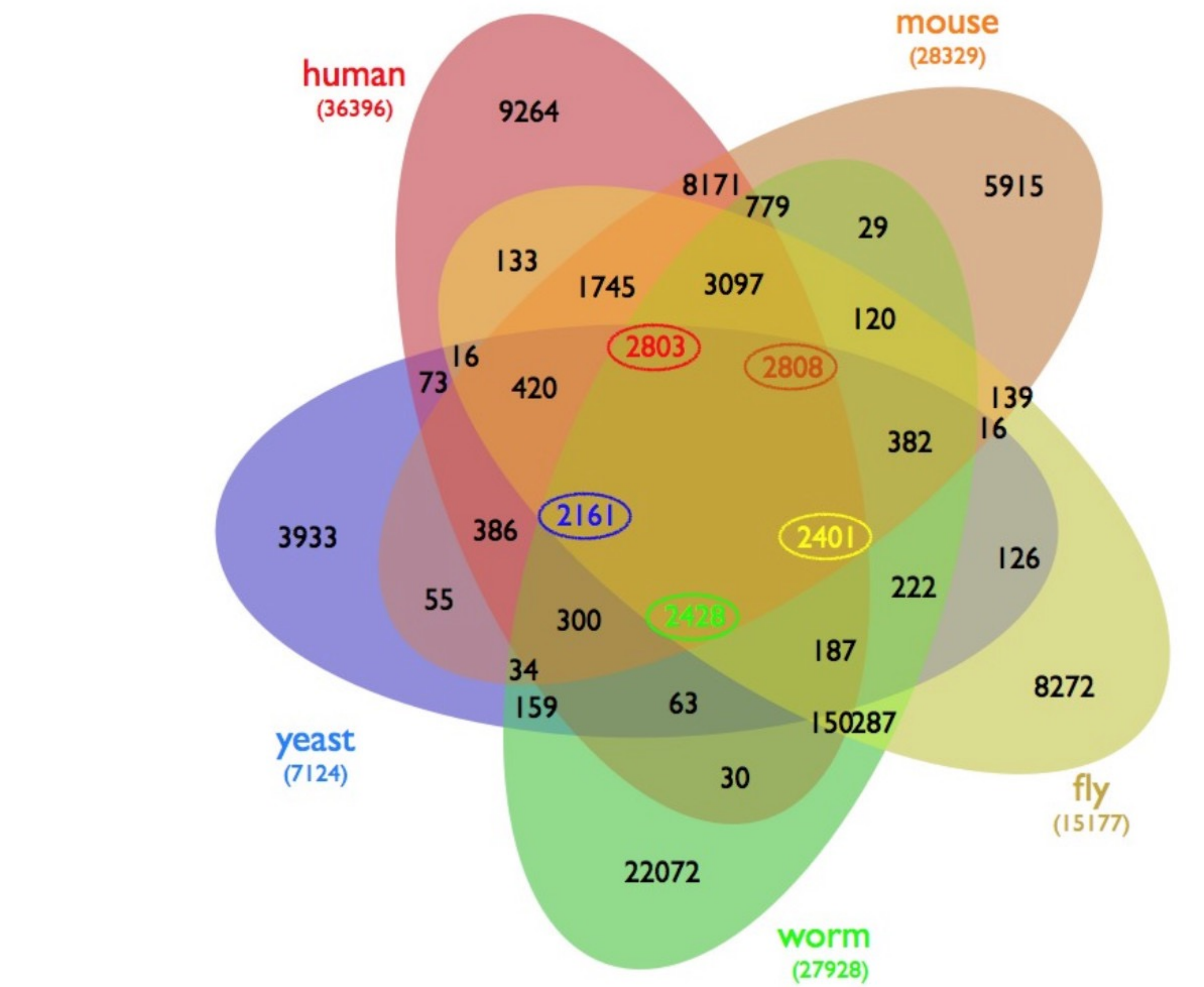
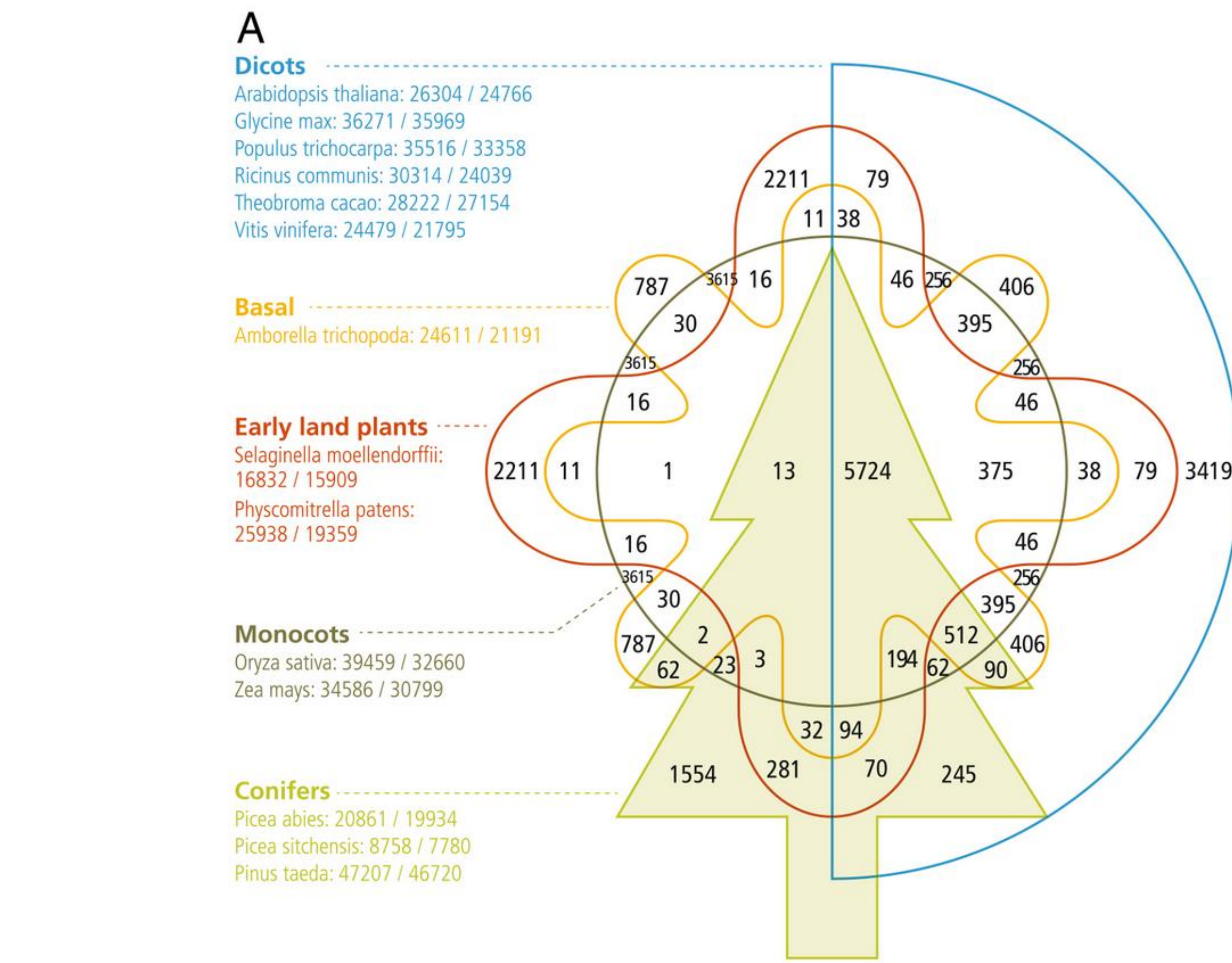
Sorghum *Sorghum bicolor*

Brome *Brachypodium distachyon*

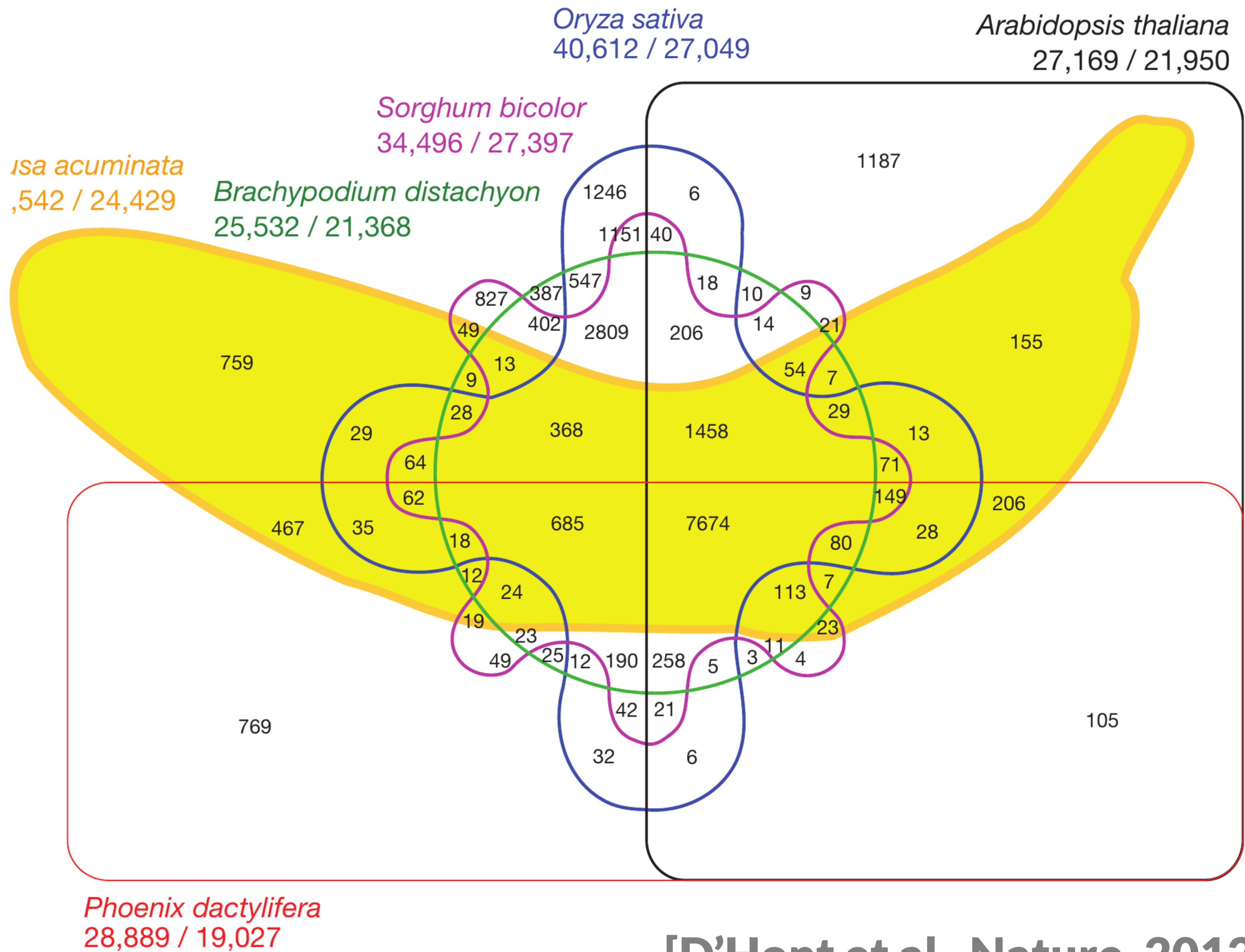


[D'Hont et al., Nature, 2012]

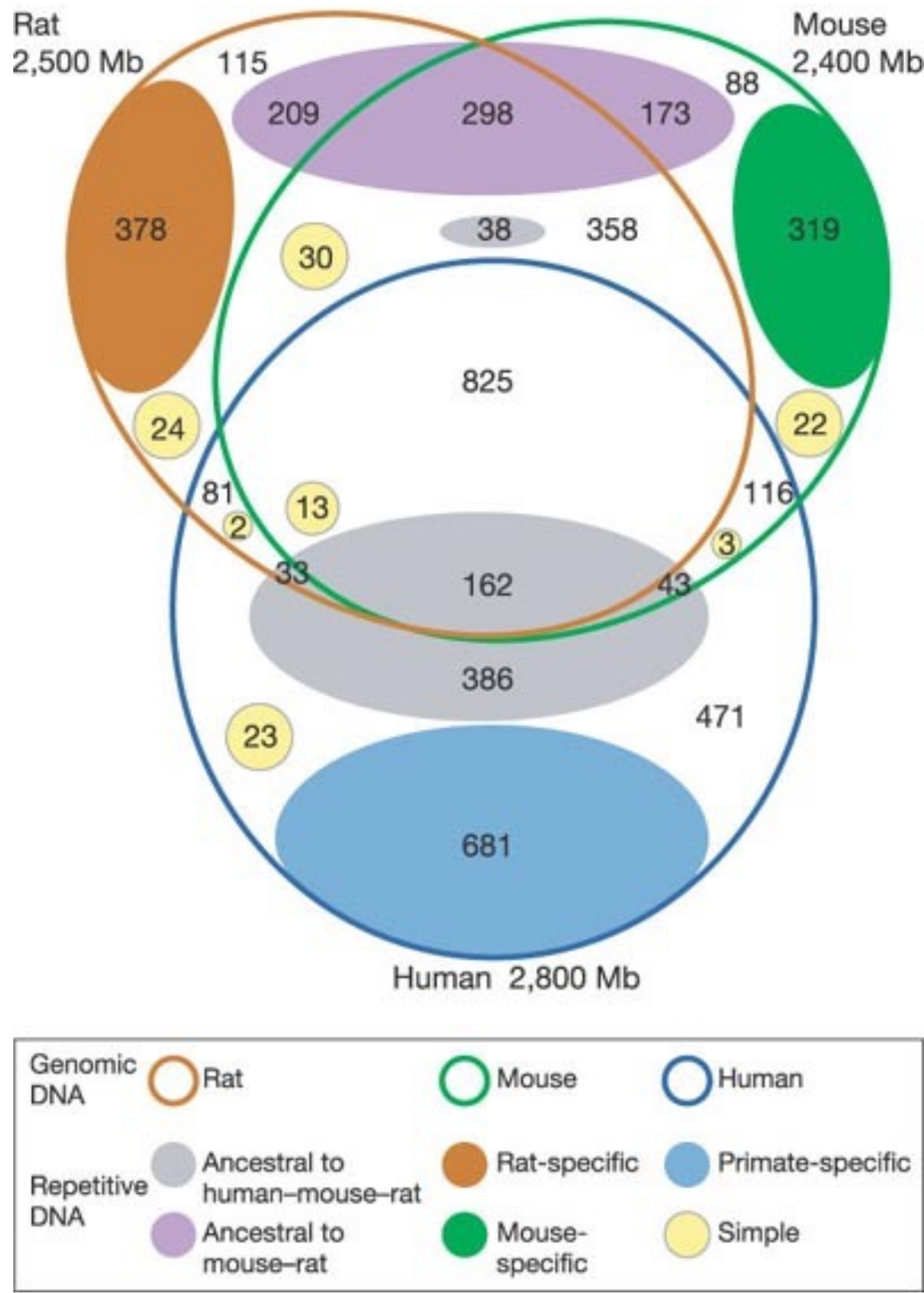
[Neale et al., BMC Genome Biology, 2014]



[Wiles et al., BMC Systems Biology]

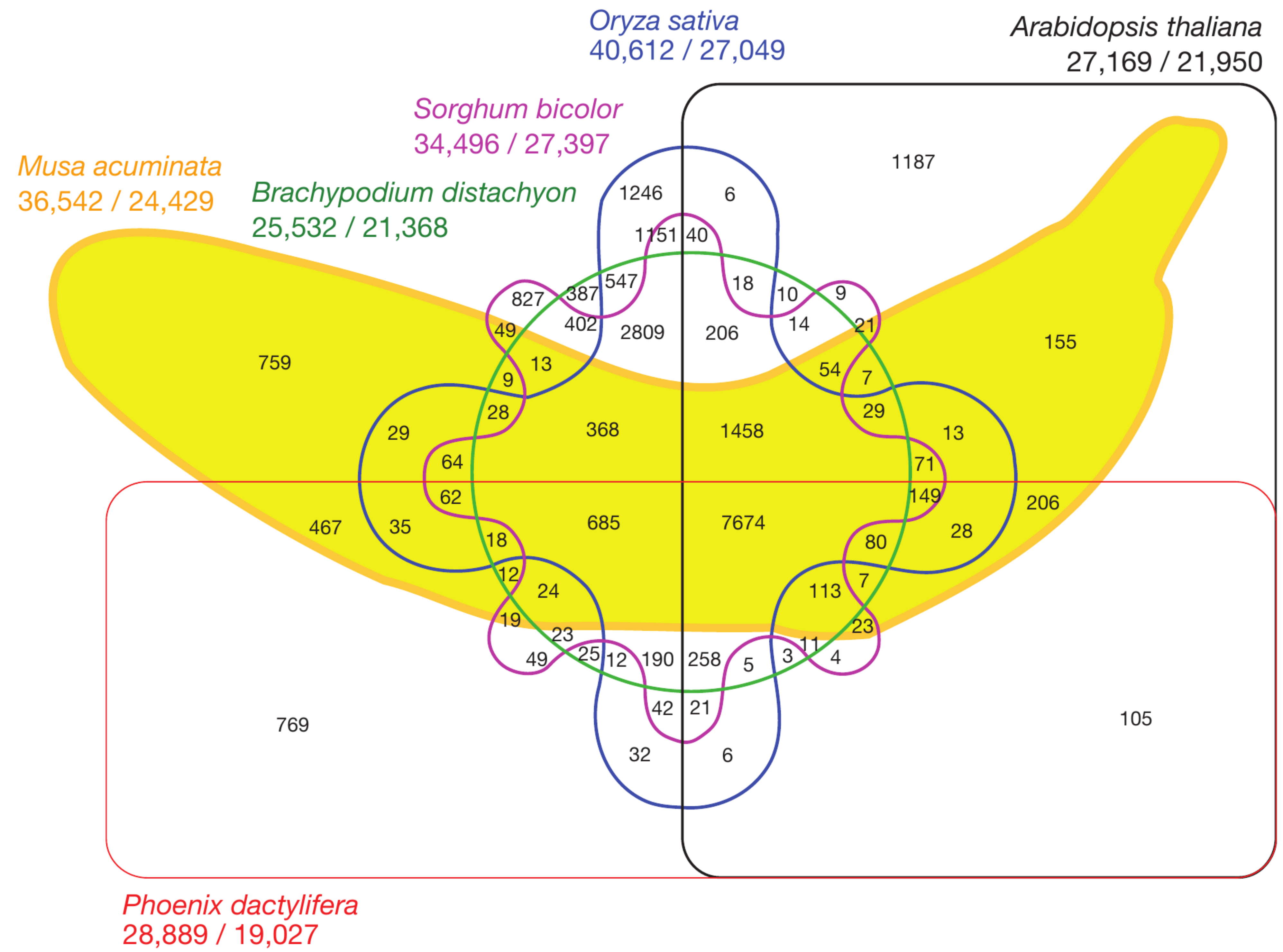


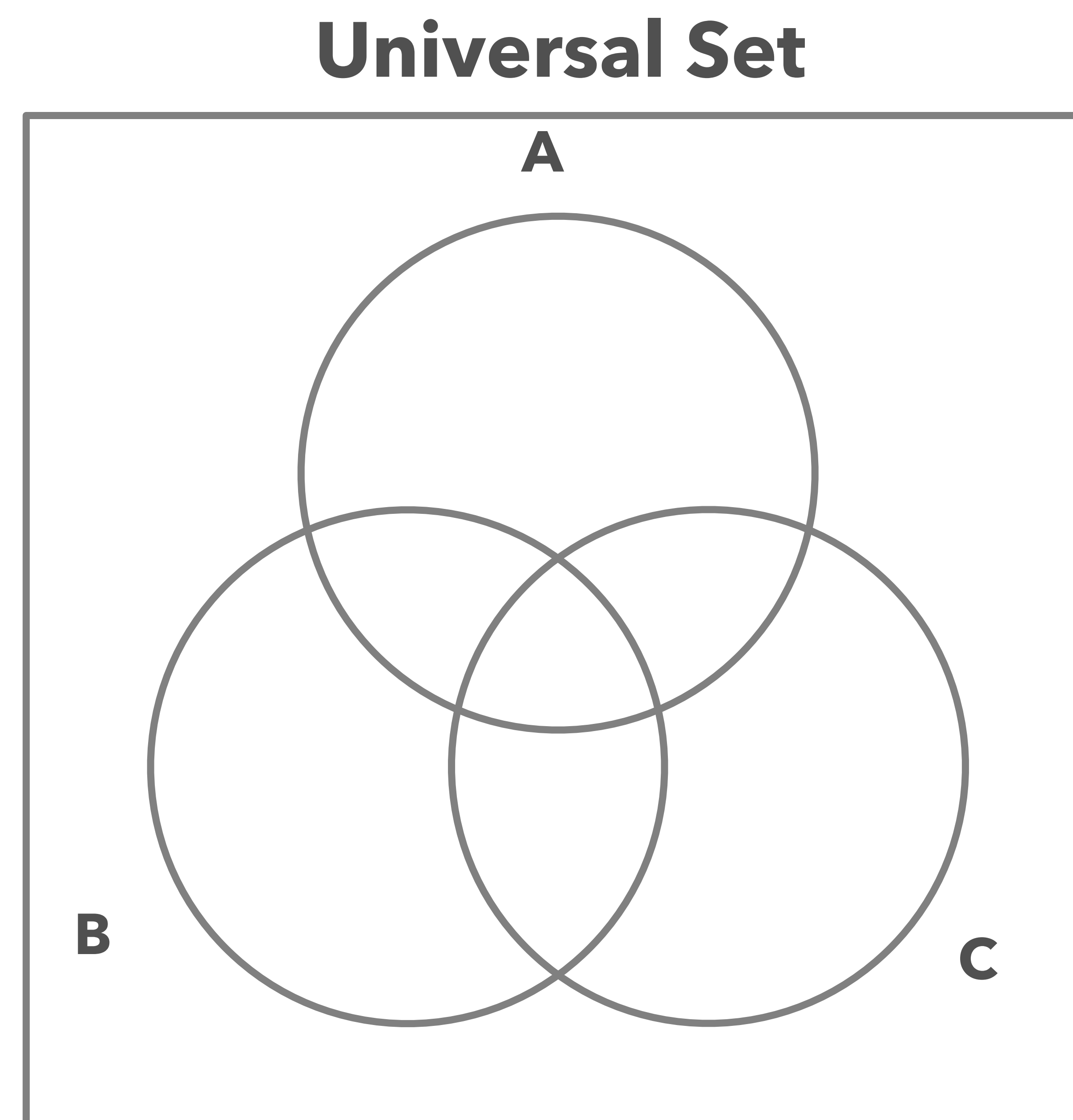
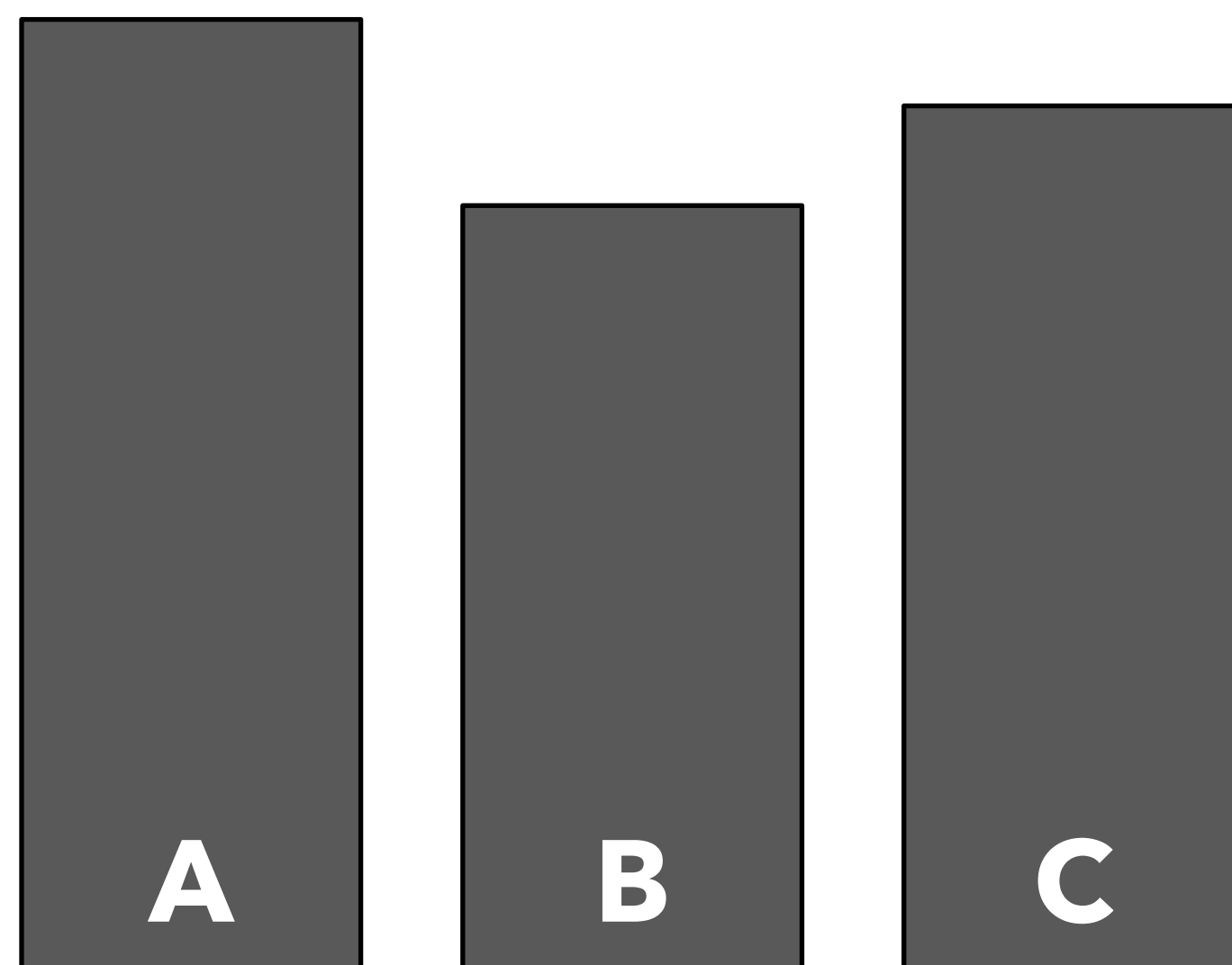
[D'Hont et al., Nature, 2012]

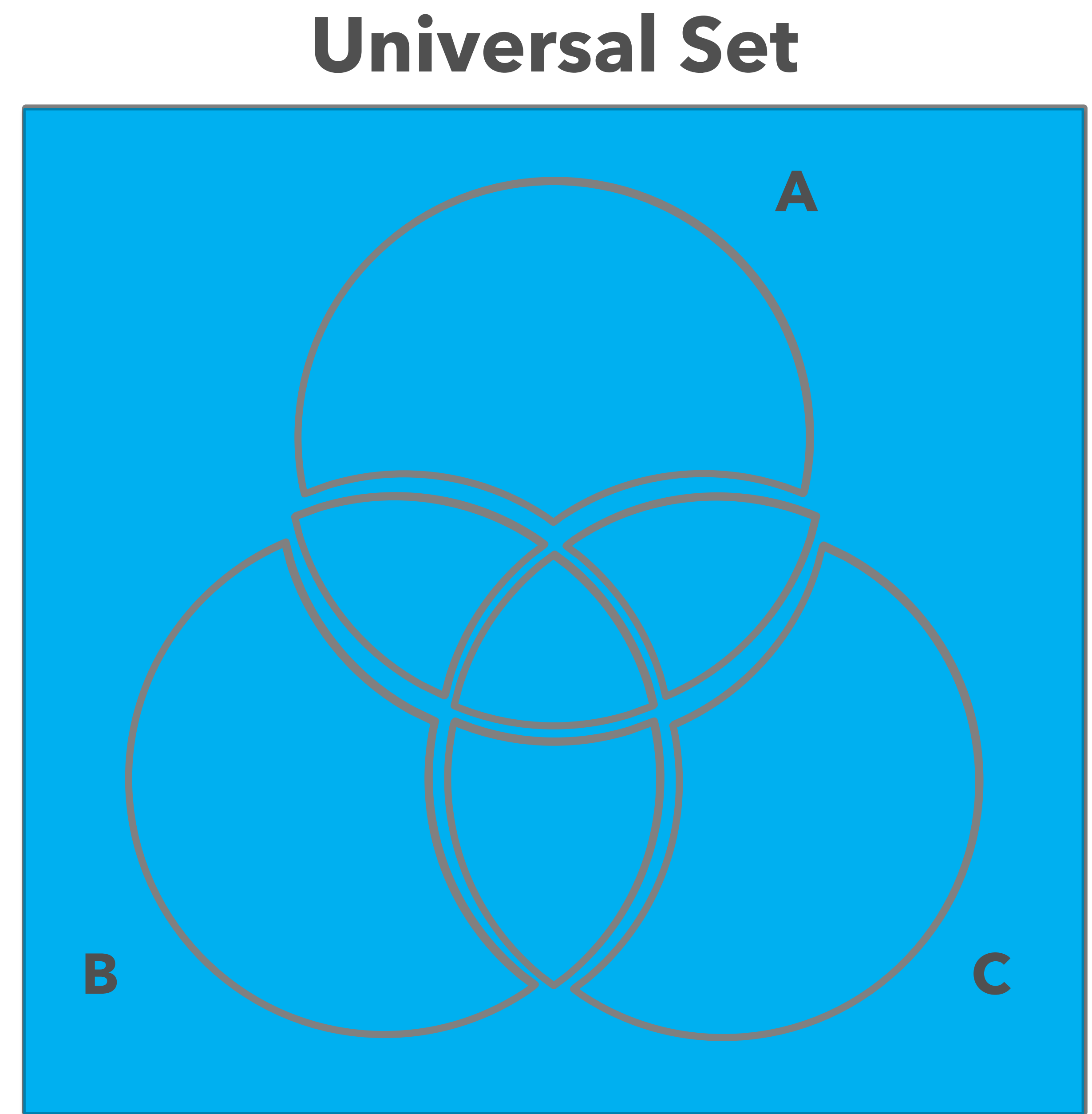
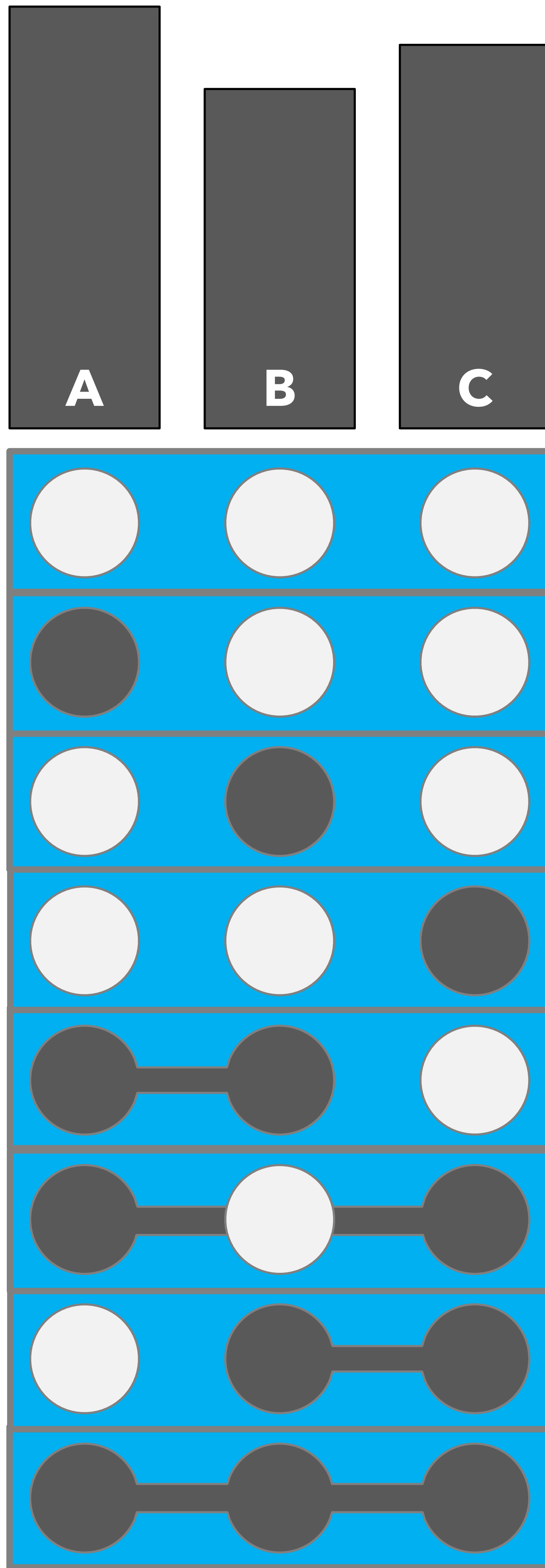


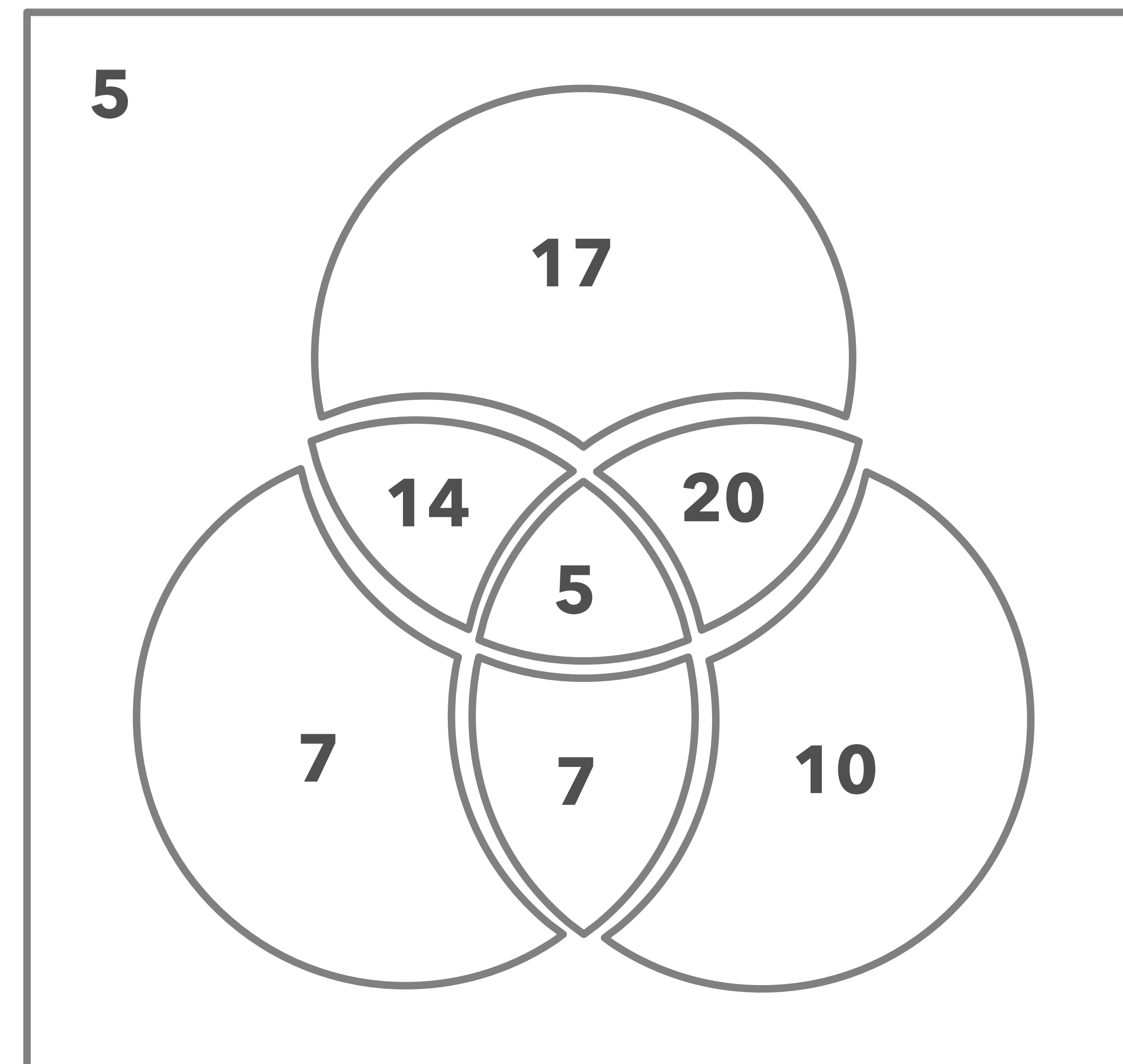
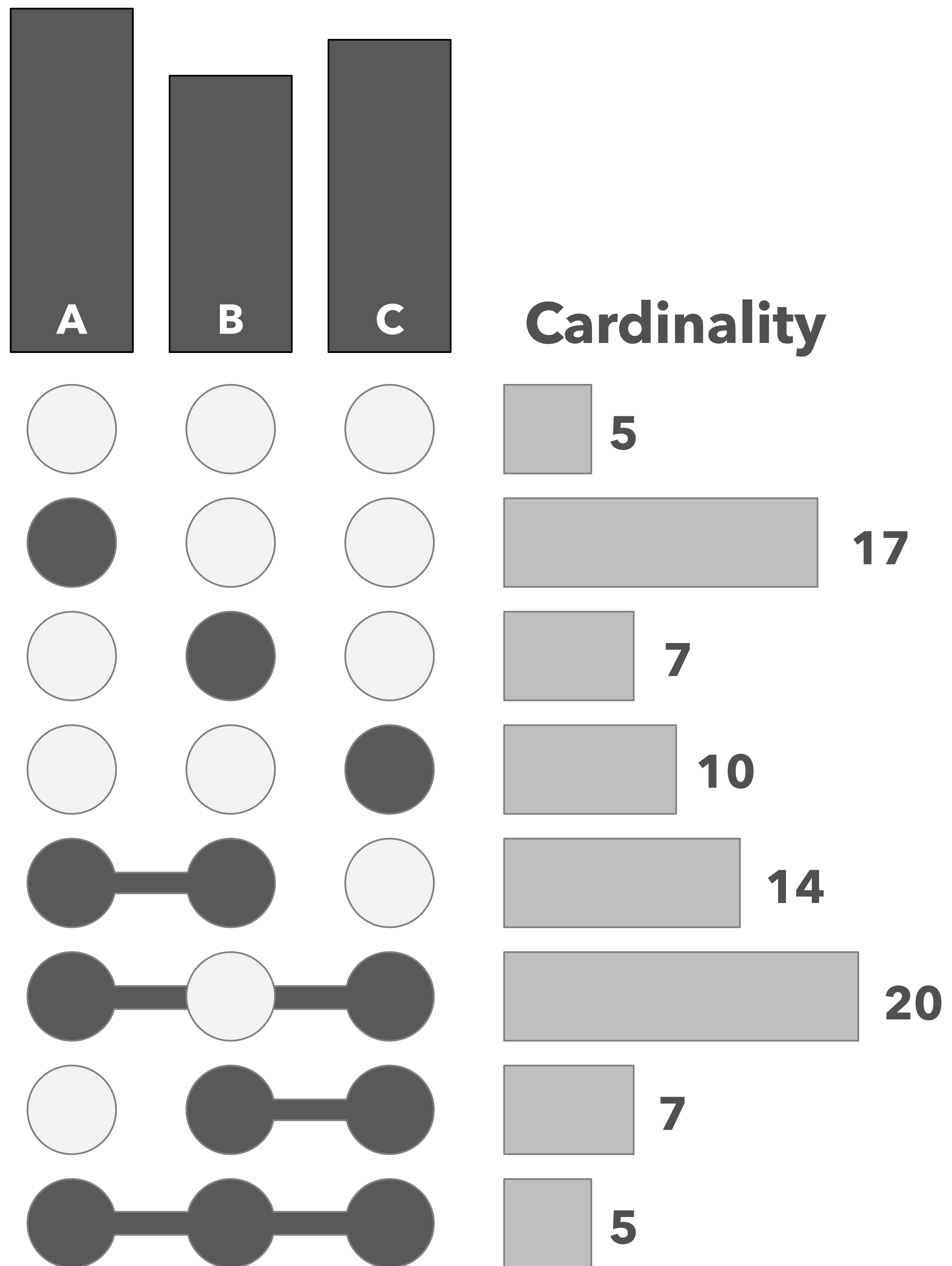
[Gibbs et al., Nature, 2004]

SO CAN WE DO BETTER?

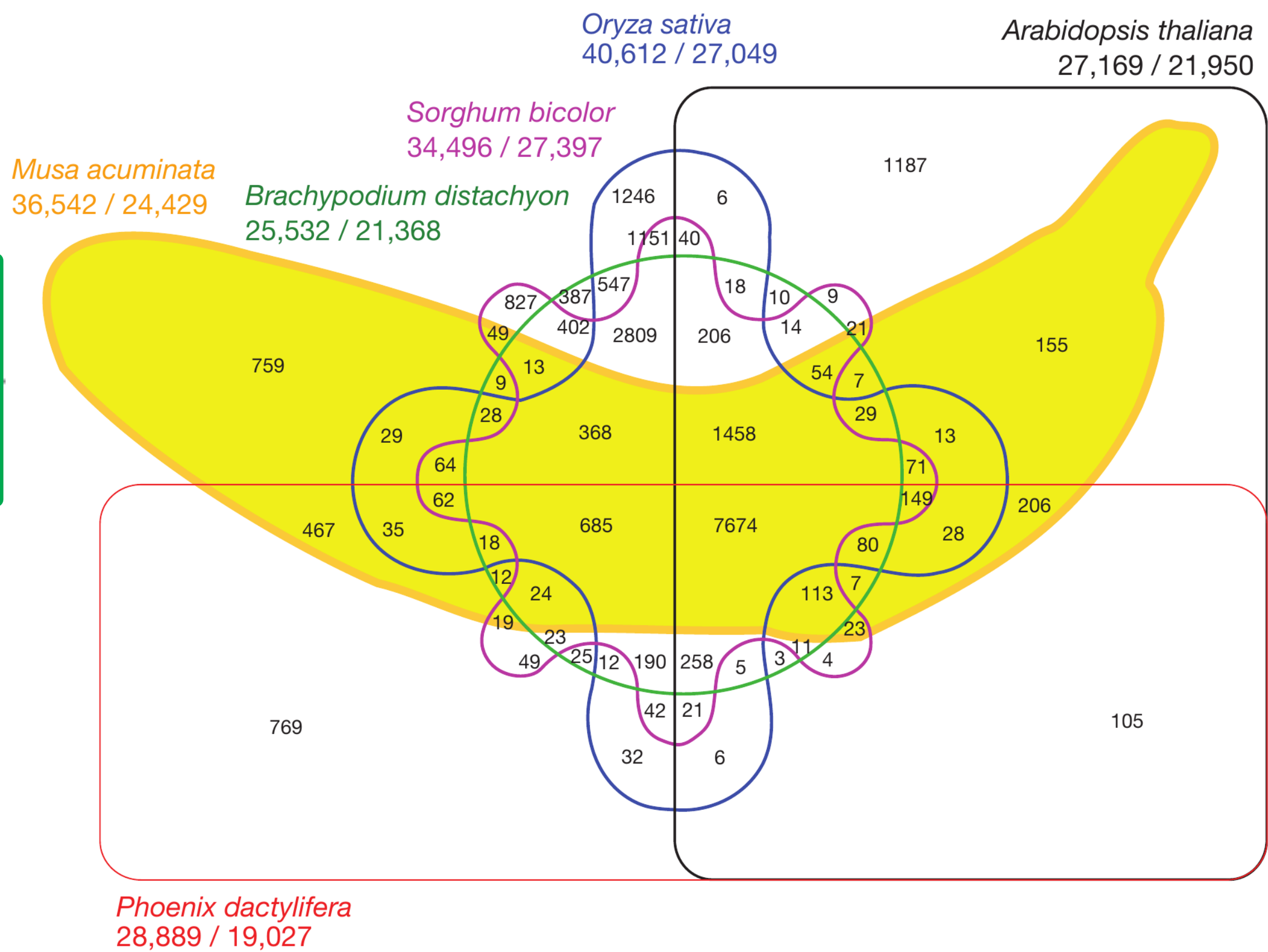
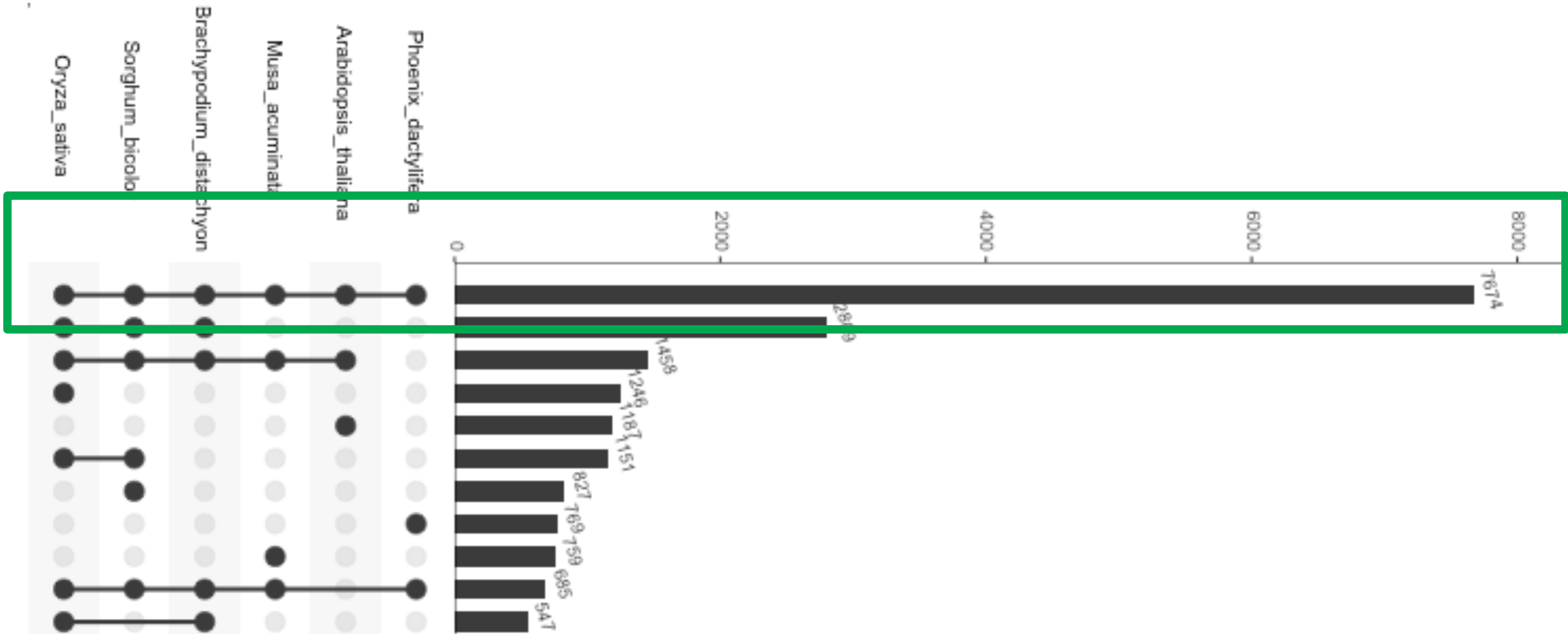






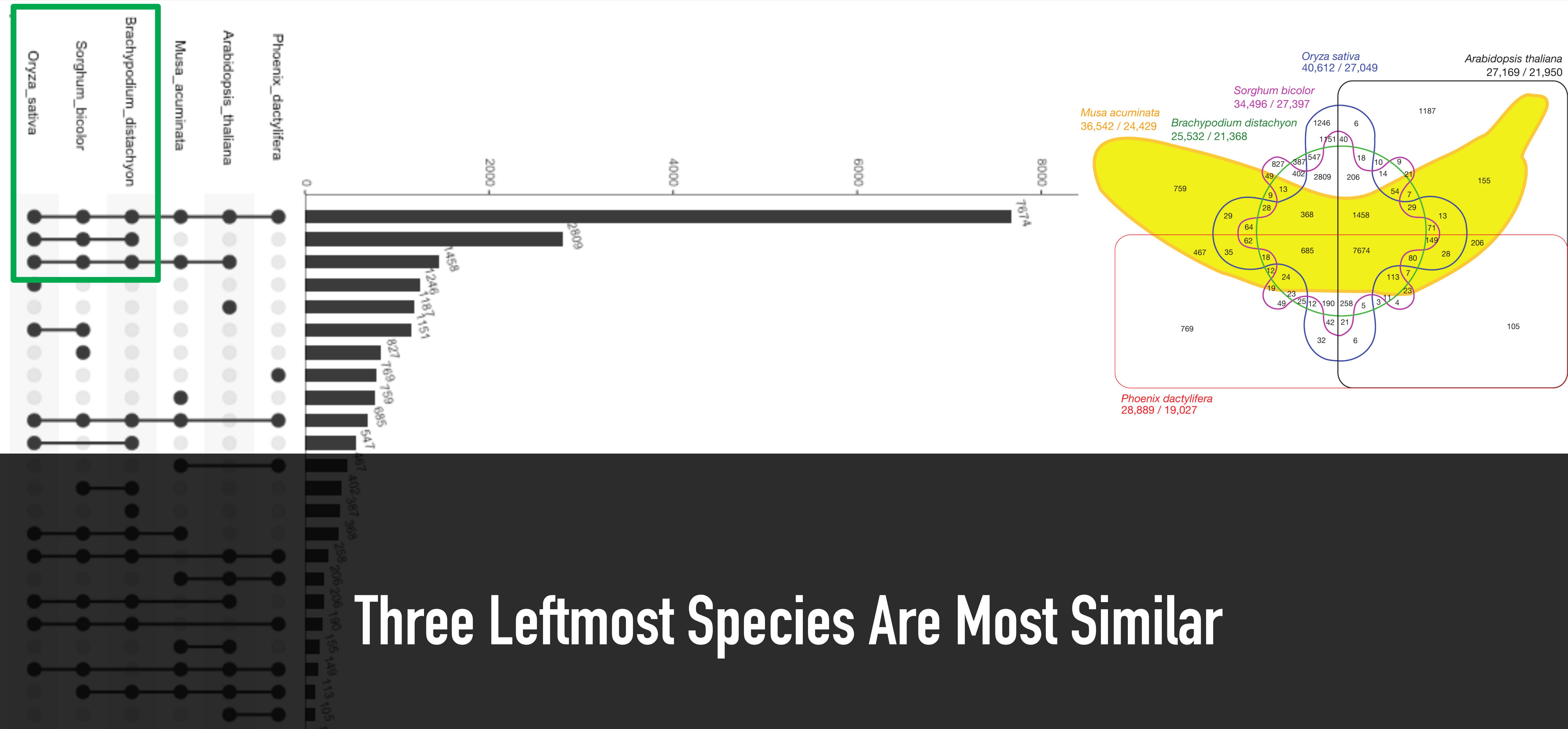


THE BANANA CHART REDESIGNED: UPSET



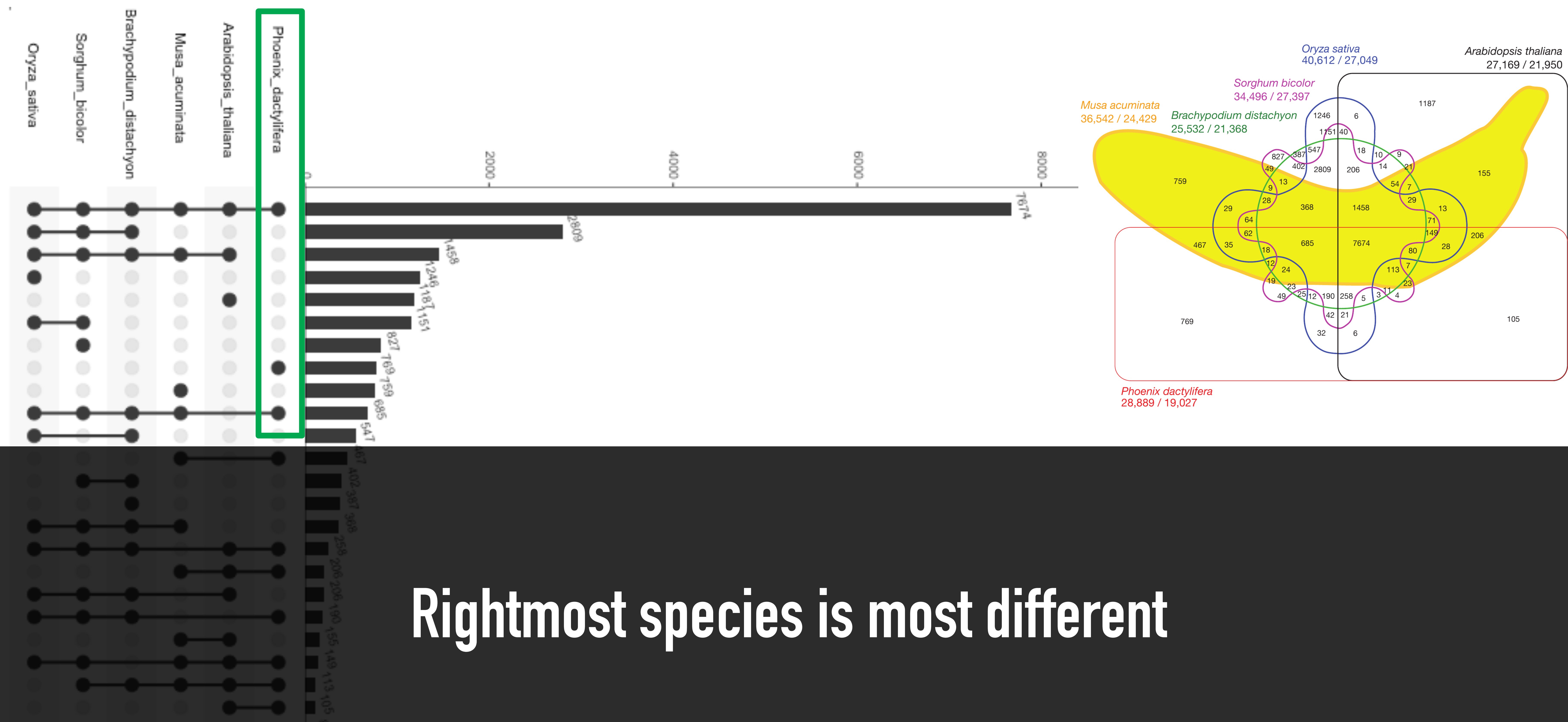
Largest Intersection Includes All Sets

THE BANANA CHART REDESIGNED: UPSET



Three Leftmost Species Are Most Similar

THE BANANA CHART REDESIGNED: UPSET



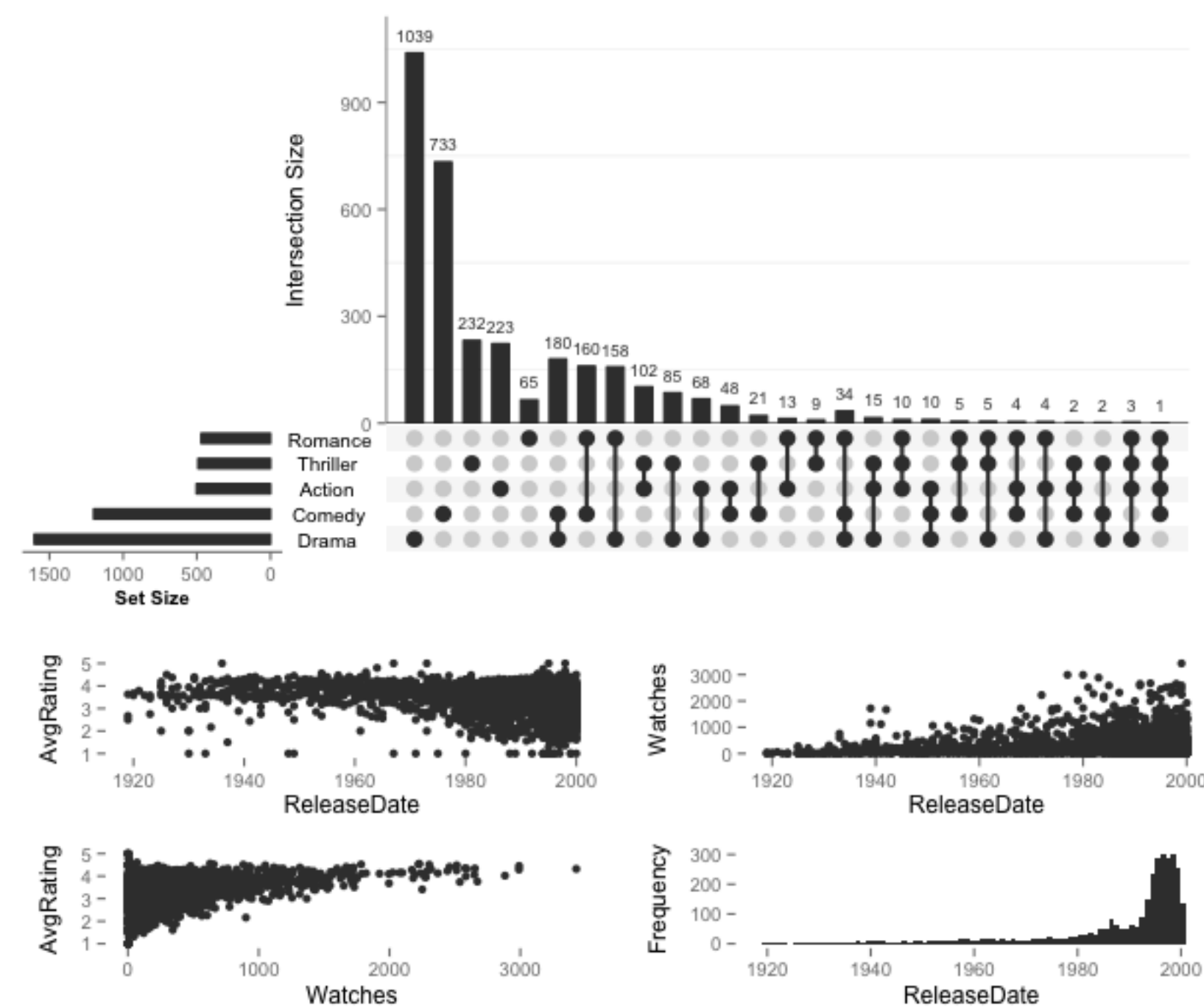
Rightmost species is most different

UPSET

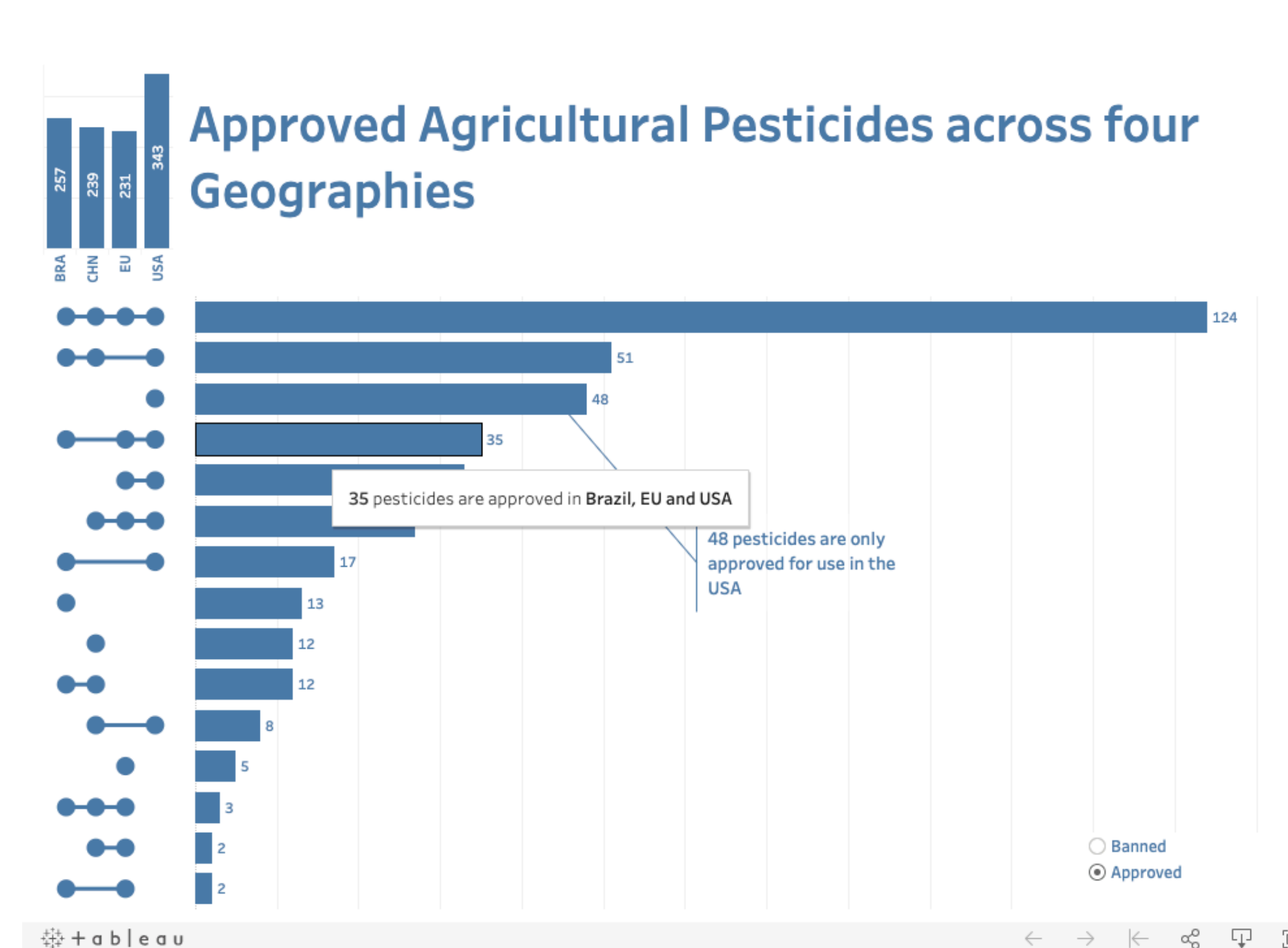
The canonical way to show set data with > 3 sets

Second-most cited VIS paper of the last decade

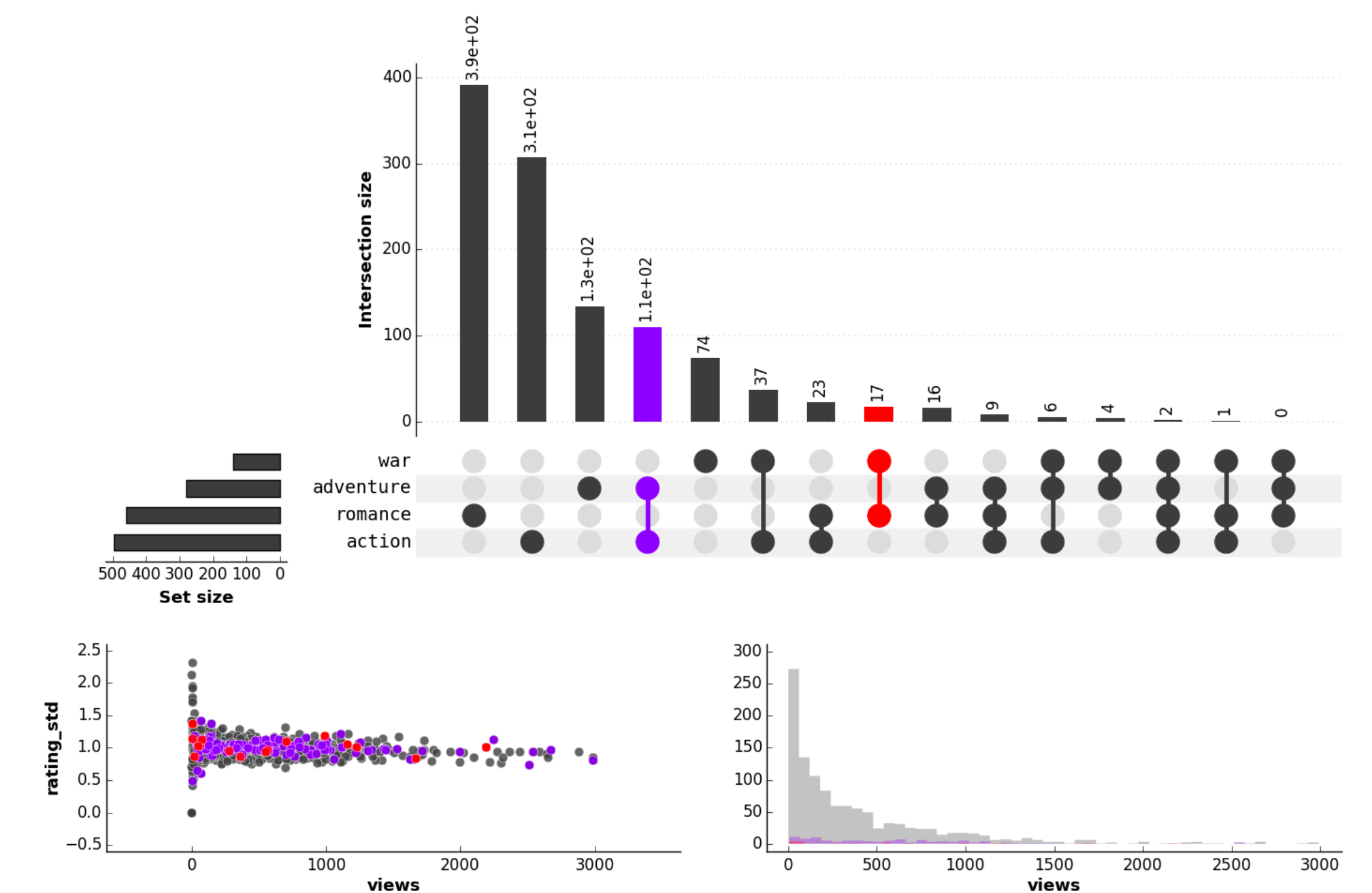
Available in various languages: <http://upset.app/>



R



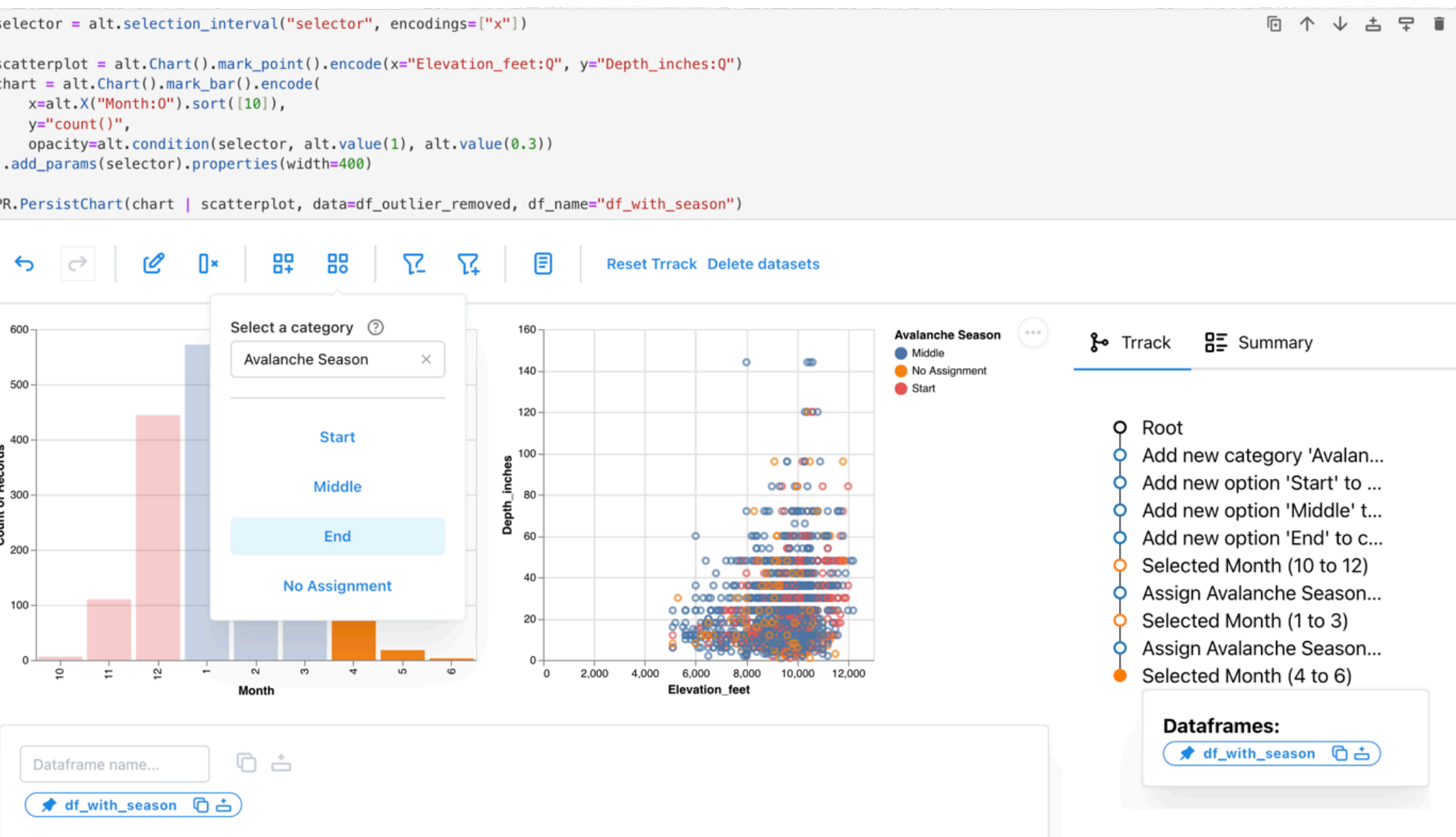
Tableau



Python

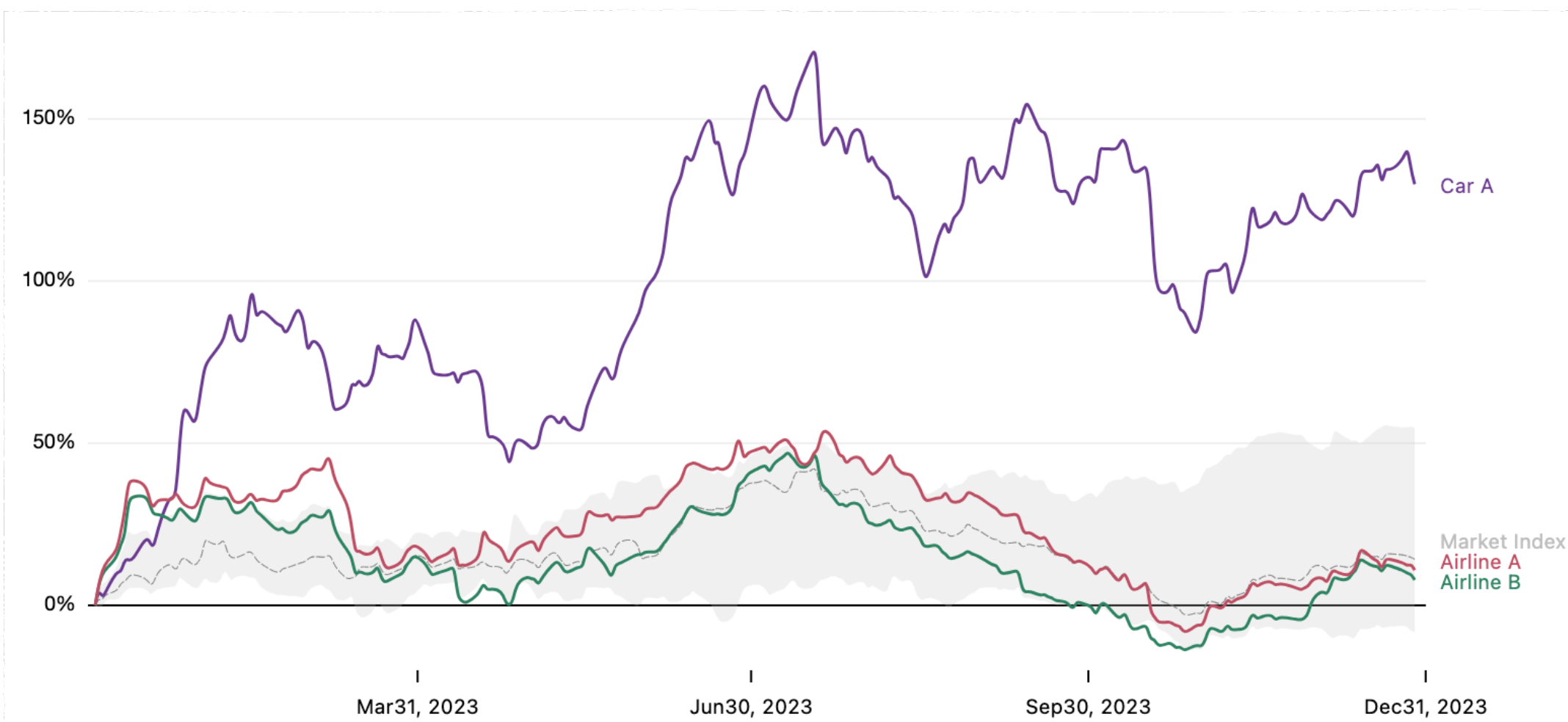
TECHNICAL CONTRIBUTIONS

Interaction in Notebooks



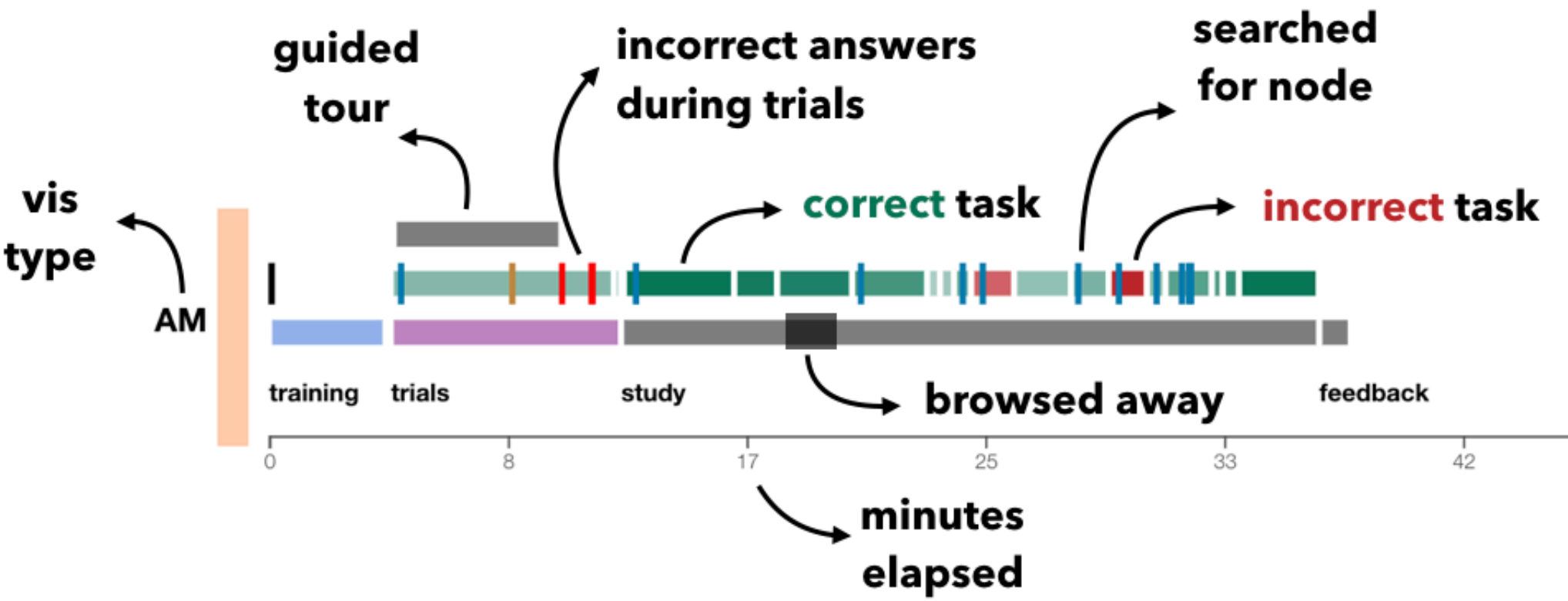
EMPIRICAL & THEORETICAL WORK

Visual Misinformation



INFRASTRUCTURE

Evaluating Complex Systems



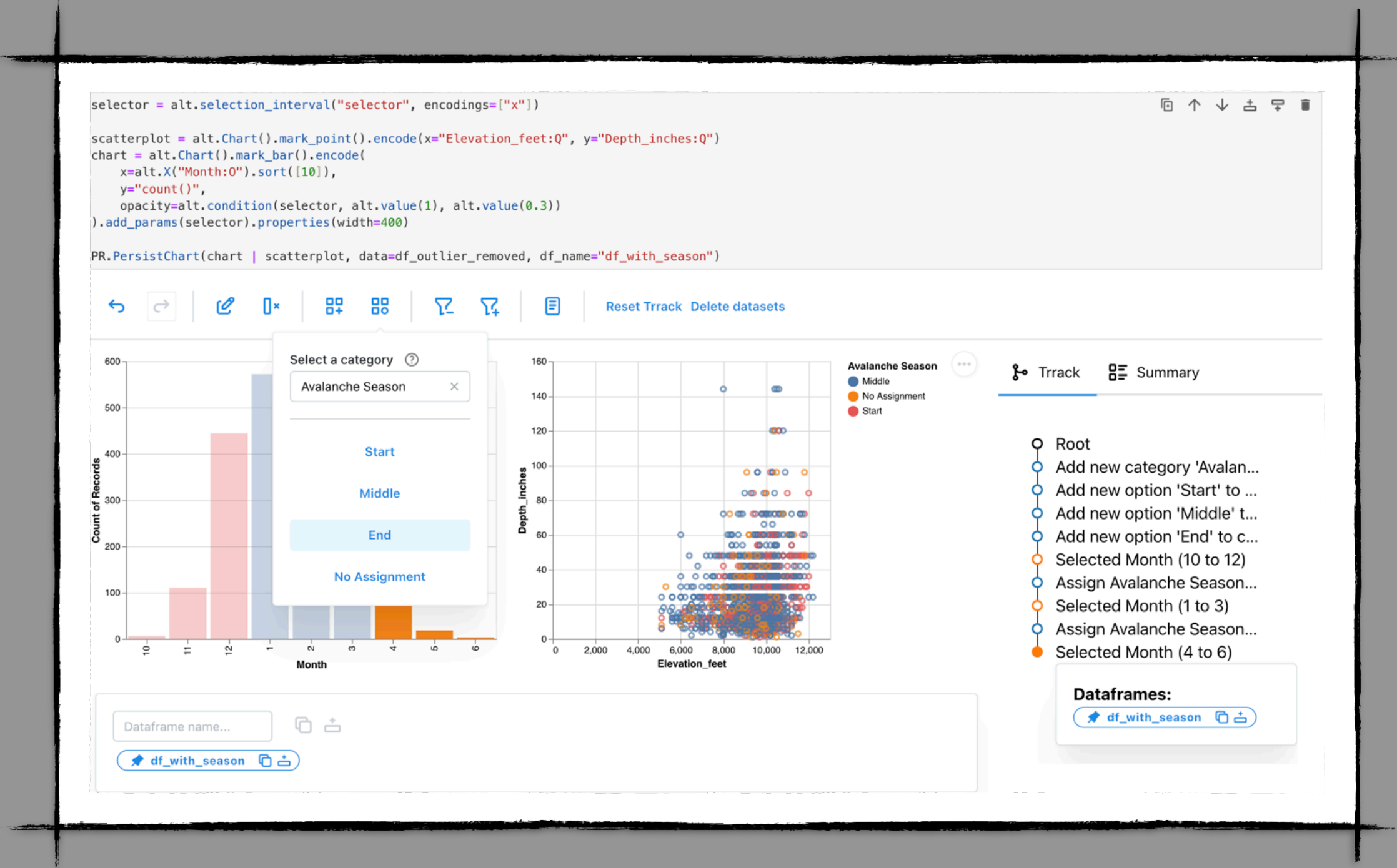
TECHNICAL CONTRIBUTIONS



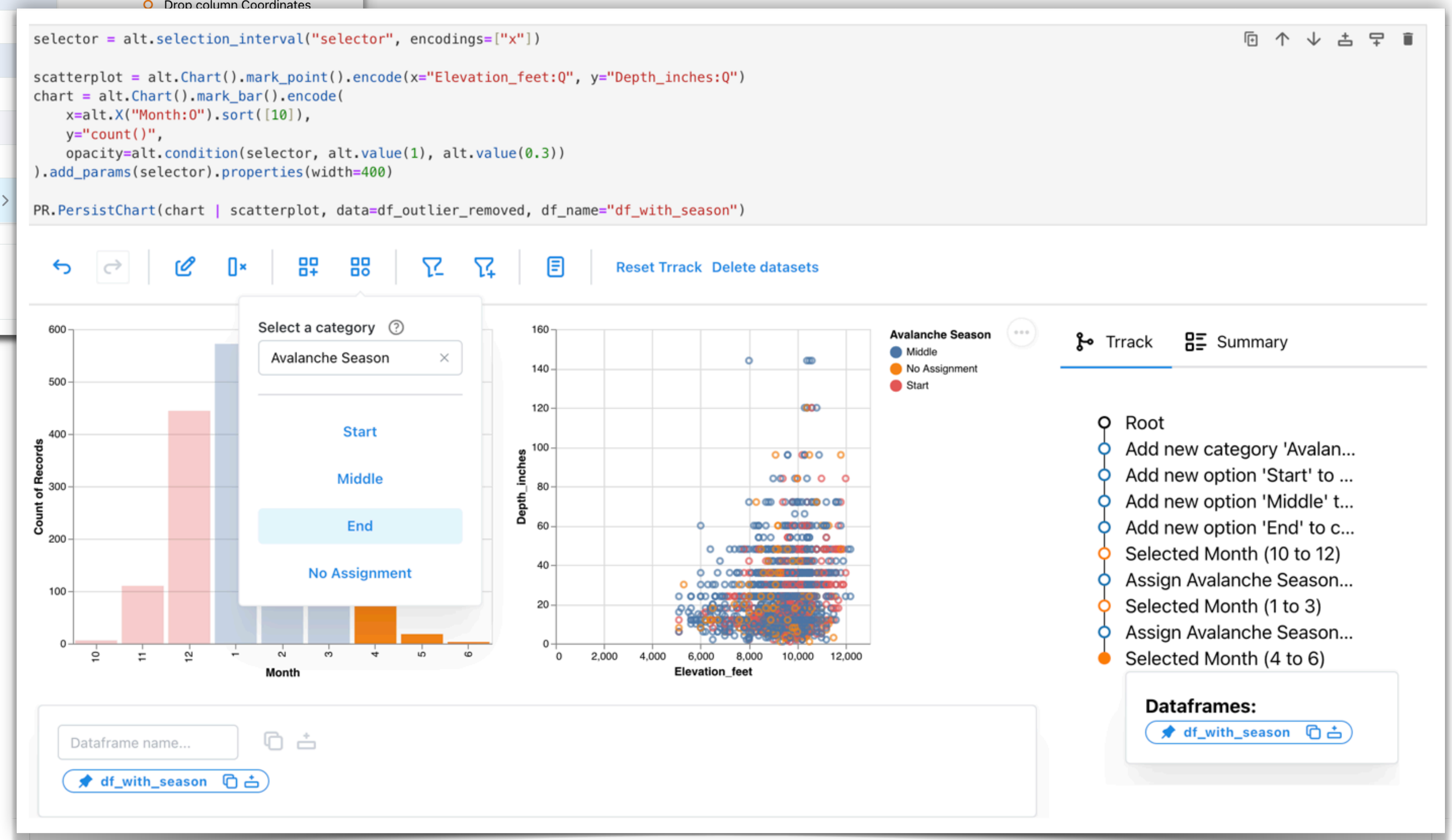
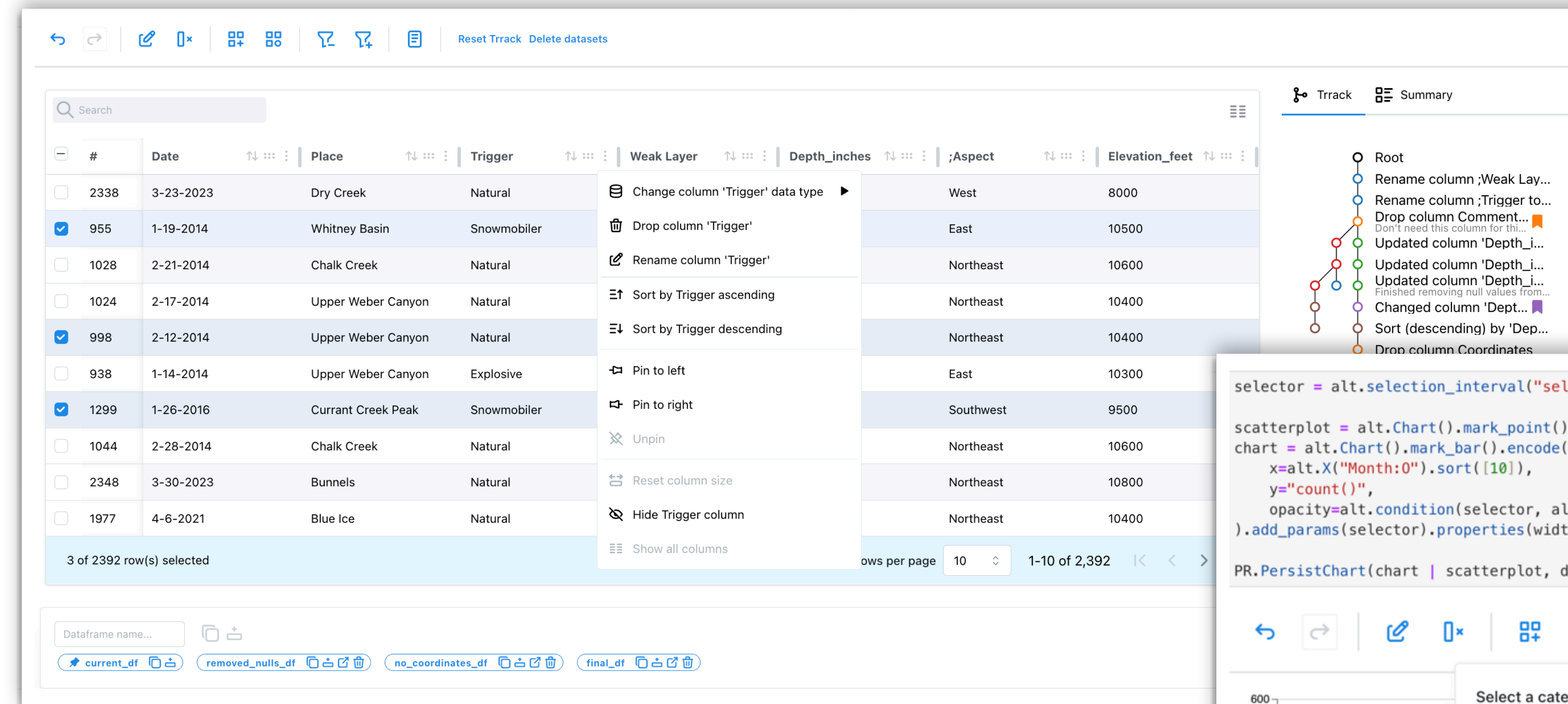
Kiran Gadhav, Zach Cutler

PERSISTENT AND REUSABLE INTERACTIONS

... IN COMPUTATIONAL NOTEBOOKS



PERSISTENT AND REUSABLE INTERACTIONS IN COMPUTATIONAL NOTEBOOKS



A Jupyter Plugin

<https://vdl.sci.utah.edu/persist/>

`pip install persist_ext`

Demo Notebook: <https://tinyurl.com/5db7nynn>

DATASET EXAMPLE: HISTORICAL AVALANCHES IN UTAH

Avalanches are a major hazard

in Utah

Utah Avalanche Center collects data about avalanches, including **where it occurred (location, elevation), how it occurred, how big it was,** etc.



WHAT IS THIS TALK ABOUT?

Supposed you're doing data analysis in Python

What's the pandas code...

- ...to **change the order of columns?**
- ...to **drop a column?**
- ...to **change the label of a column?**

Nothing here is hard, but it's annoying.

PERSIST MAKES THIS EASY

[4]: PR.PersistTable(avalanches, df_name="avalanches")

Reset Ttrack

Delete datasets

Search

<input type="checkbox"/>	#	;Region	Month	Day	Year	;Trigger	;Weak Layer
<input type="checkbox"/>	1	Salt Lake	11	9	2012	Snowboarder	New Snow/Old Snow
<input type="checkbox"/>	2	Salt Lake	11	11	2012	Skier	New Snow/Old Snow
<input type="checkbox"/>	3	Salt Lake	11	11	2012	Skier	Facets
<input type="checkbox"/>	4	Salt Lake	11	11	2012	Skier	New Snow
<input type="checkbox"/>	5	Salt Lake	11	11	2012	Skier	Facets
<input type="checkbox"/>	6	Salt Lake	11	10	2012	Skier	New Snow/Old Snow
<input type="checkbox"/>	7	Salt Lake	11	12	2012	Skier	Facets
<input type="checkbox"/>	8	Salt Lake	12	8	2012	Skier	Facets
<input type="checkbox"/>	9	Salt Lake	12	9	2012	Skier	Facets
<input type="checkbox"/>	10	Salt Lake	12	10	2012	Skier	Facets

Rows per page

10

1-10 of 2,392

Ttrack

Summary

● Root

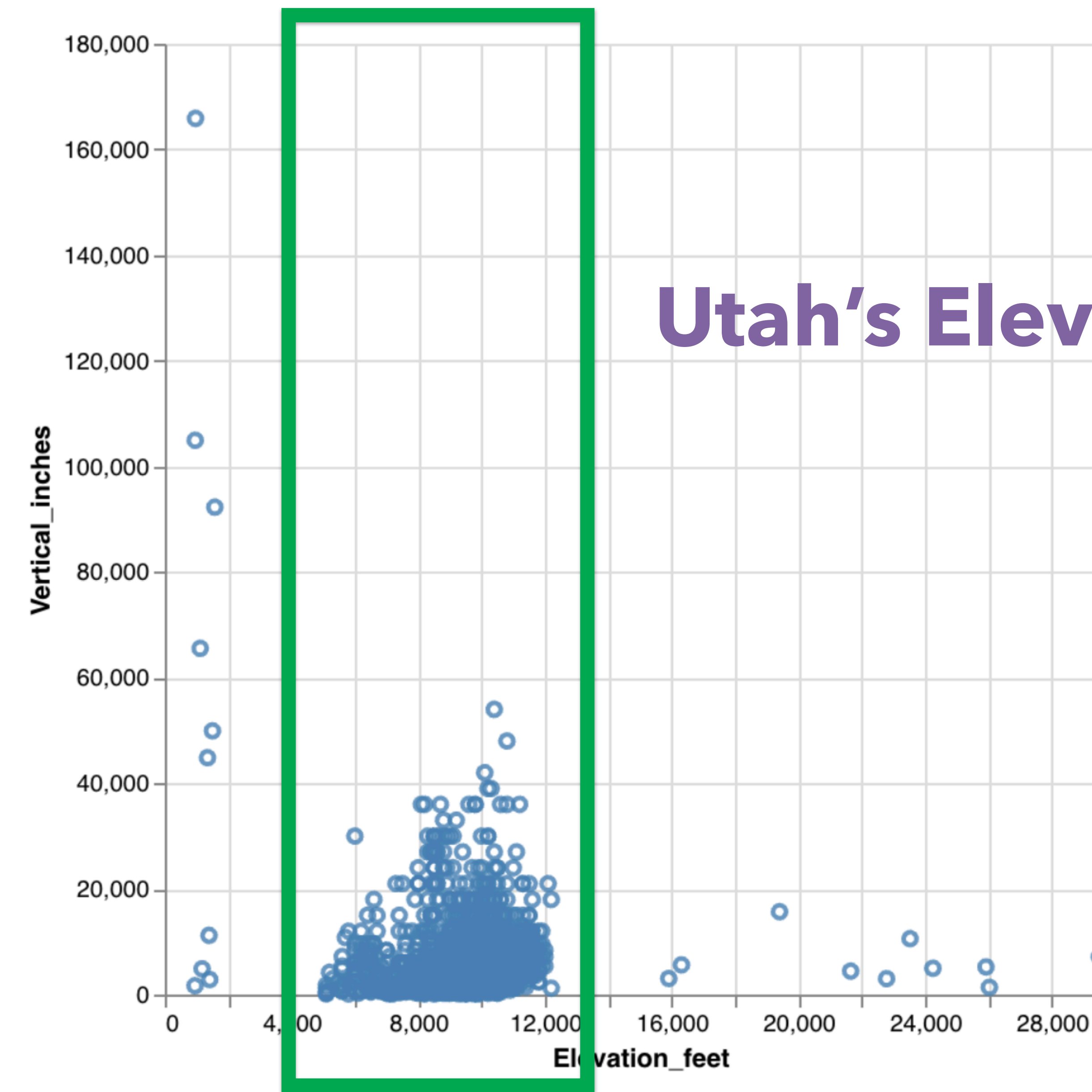
Dataframe name...

avalanches

WHAT IS THIS TALK ABOUT?

Have you ever plotted something and wished you could just “fix” things as you spot them?

How deep (big) the avalanche was



Utah's Elevation Range

Elevation where the avalanche occurred

PERSIST MAKES THIS EASY

```
[6]: PR.plot.scatterplot(avalanches, "Elevation_feet:Q", "Vertical_inches:Q", df_name="avalanches")
```

↶

↷

✎

×

⊞

⊞

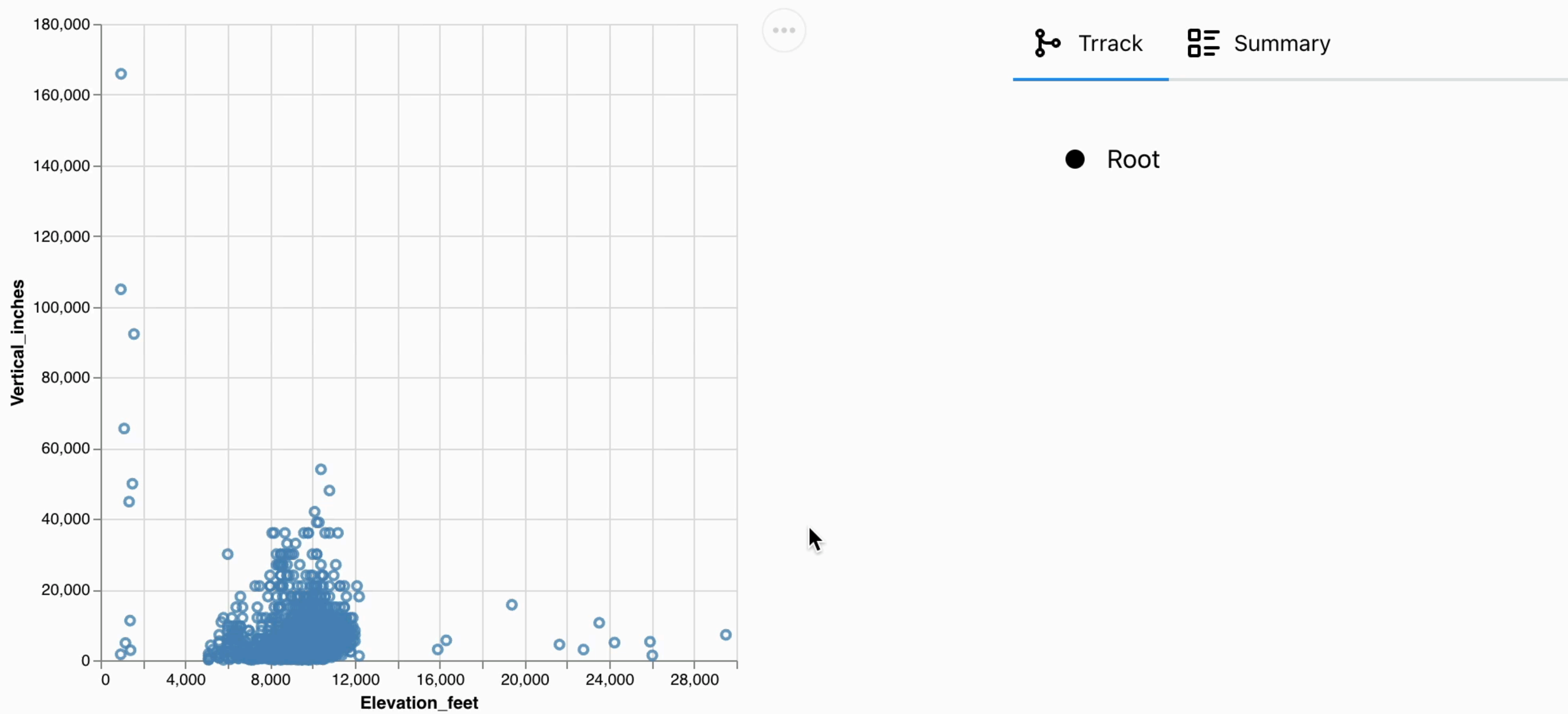
⌵

⌵

☰

Reset Ttrack

Delete datasets



SO WHAT'S SPECIAL HERE?

Lots of **vis tools** support these operations

Most **data wrangling happens in code**: it's just more powerful

Opportunity: bring **interactive operations to code**!

Persist works **INSIDE** your Jupyter Notebook

BRIDGING BETWEEN DATA ANALYSIS MODALITIES

**BRIDGING BETWEEN
DATA ANALYSIS
MODALITIES**

What are Modalities?

1. Interactive Vis

2. Code

INTERACTIVE VISUALIZATION: BENEFITS

Intuitive

Easy to use

Uses human perceptual capabilities



INTERACTIVE VISUALIZATION: DOWNSIDES

Limited Expressivity

Some operations are difficult

e.g., conditional queries..

Not reusable

need to redo analysis when data changes

Not reproducible



CODE: BENEFITS

Flexible and powerful

you basically can do anything

Reusable

if your data changes, re-run

Reproducible

everything is documented

```
1 # Keep this cell
2 avy_df = pd.read_csv('./avalanches.csv')
3
4 # Remove NaN coordinates
5 avy_df = avy_df[avy_df['Coordinates']!=avy_df['Coordinates']]
6
7 # Split into latitude & longitude
8 avy_df[['lat', 'lon']] = avy_df['Coordinates'].str.split(',', expand=True)
9
10 # Remove values outside of Utah bounds
11 avy_df = avy_df[(36 < avy_df['lat'].astype('float')) & (avy_df['lat'].astype('float') < 42)]
12 avy_df = avy_df[(-114 < avy_df['lon'].astype('float')) & (avy_df['lon'].astype('float') < -108)]
13
14 # Keep columns we need
15 avy_df = avy_df[['Date', 'Region', 'Trigger', 'lat', 'lon']]
16 avy_df.head()
```


CODE / SCRIPTING: DOWNSIDES

It's hard

requires extensive training

reading documentation

not discoverable

It's time consuming

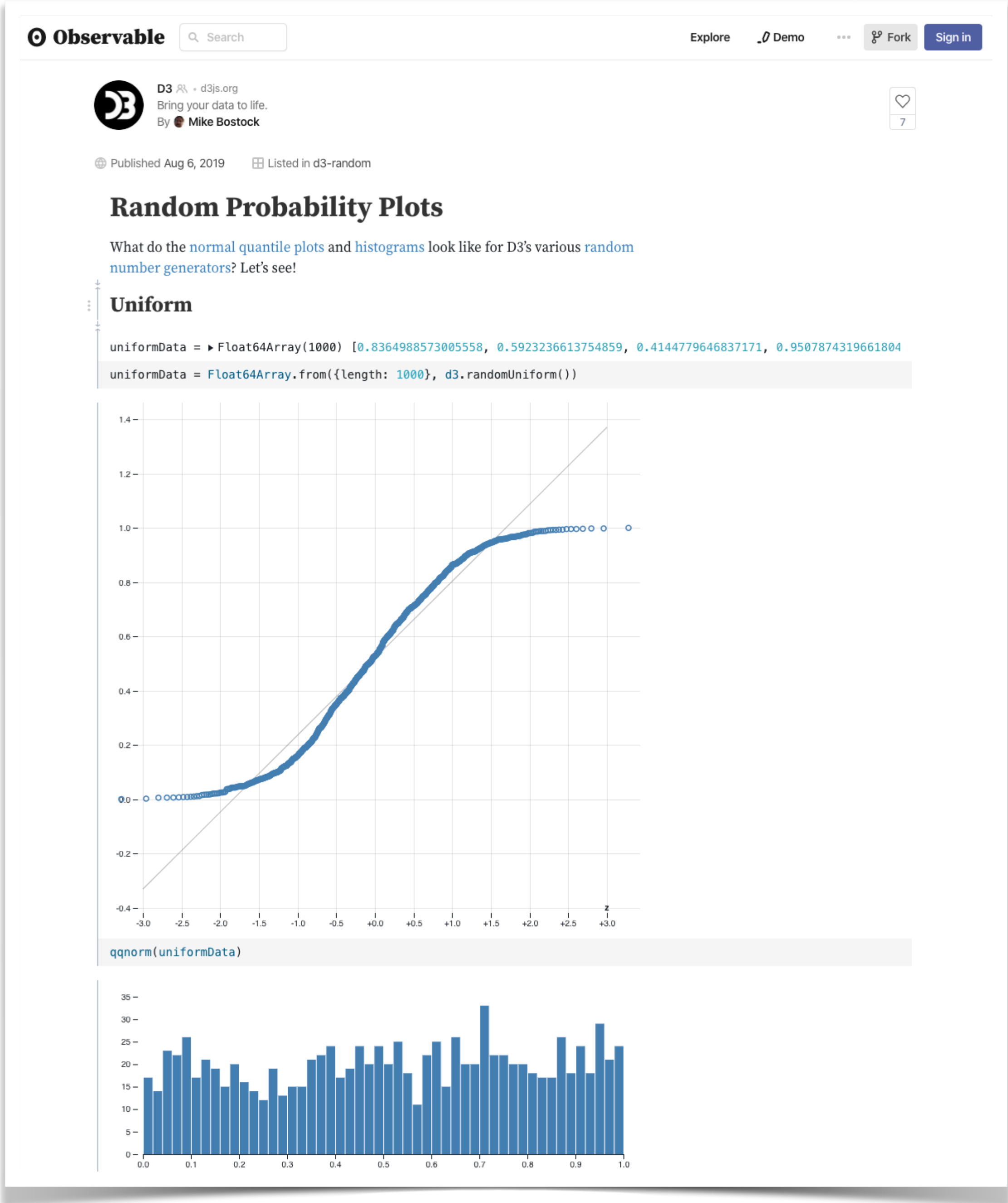
even simple things require effort

Some operations are difficult

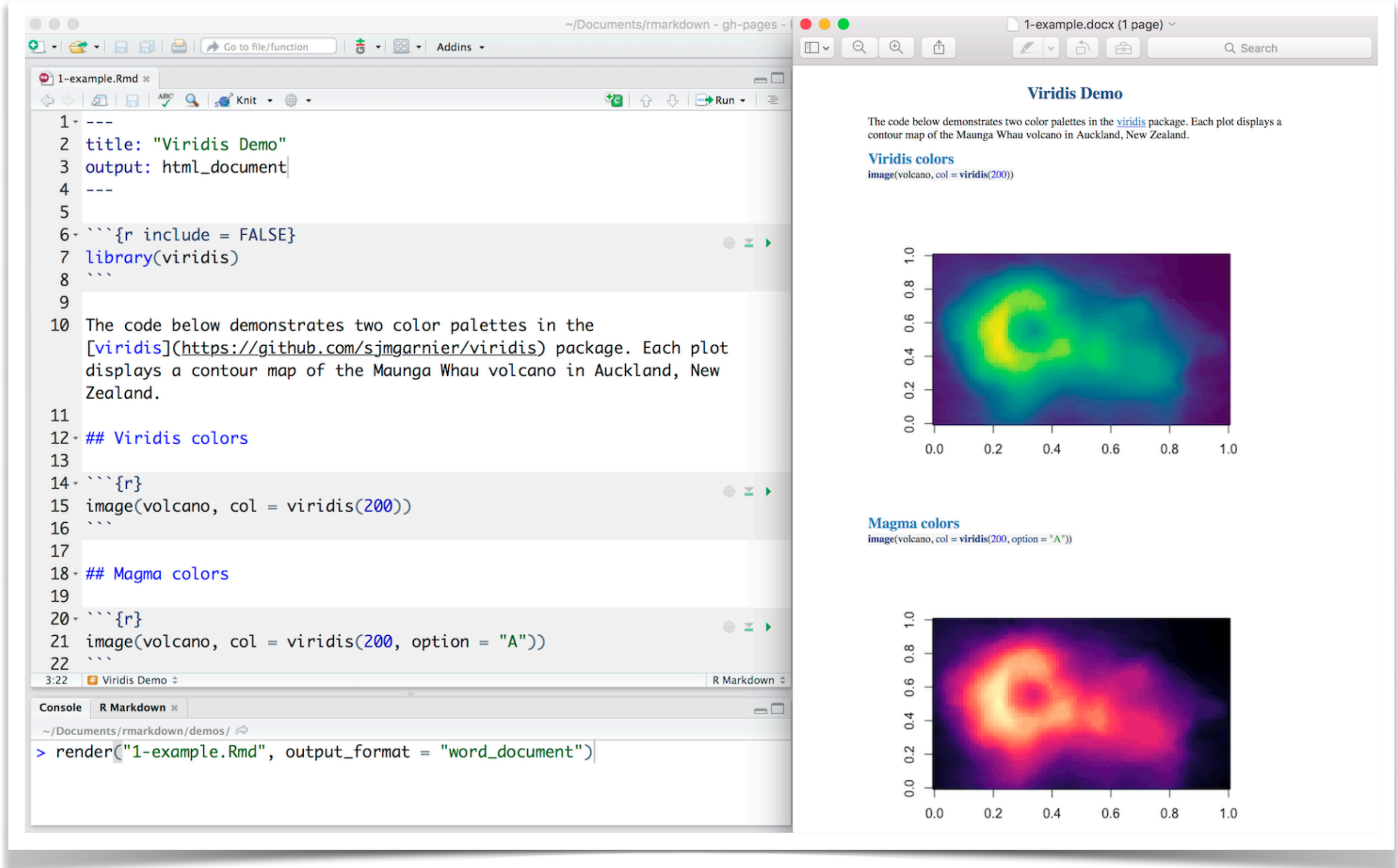
e.g., labeling data points

```
1 # Keep this cell
2 avy_df = pd.read_csv('./avalanches.csv')
3
4 # Remove NaN coordinates
5 avy_df = avy_df[avy_df['Coordinates']!=avy_df['Coordinates']]
6
7 # Split into latitude & longitude
8 avy_df[['lat', 'lon']] = avy_df['Coordinates'].str.split(',', expand=True)
9
10 # Remove values outside of Utah bounds
11 avy_df = avy_df[(36 < avy_df['lat'].astype('float')) & (avy_df['lat'].astype('float') < 42)]
12 avy_df = avy_df[(-114 < avy_df['lon'].astype('float')) & (avy_df['lon'].astype('float') < -108)]
13
14 # Keep columns we need
15 avy_df = avy_df[['Date', 'Region', 'Trigger', 'lat', 'lon']]
16 avy_df.head()
```

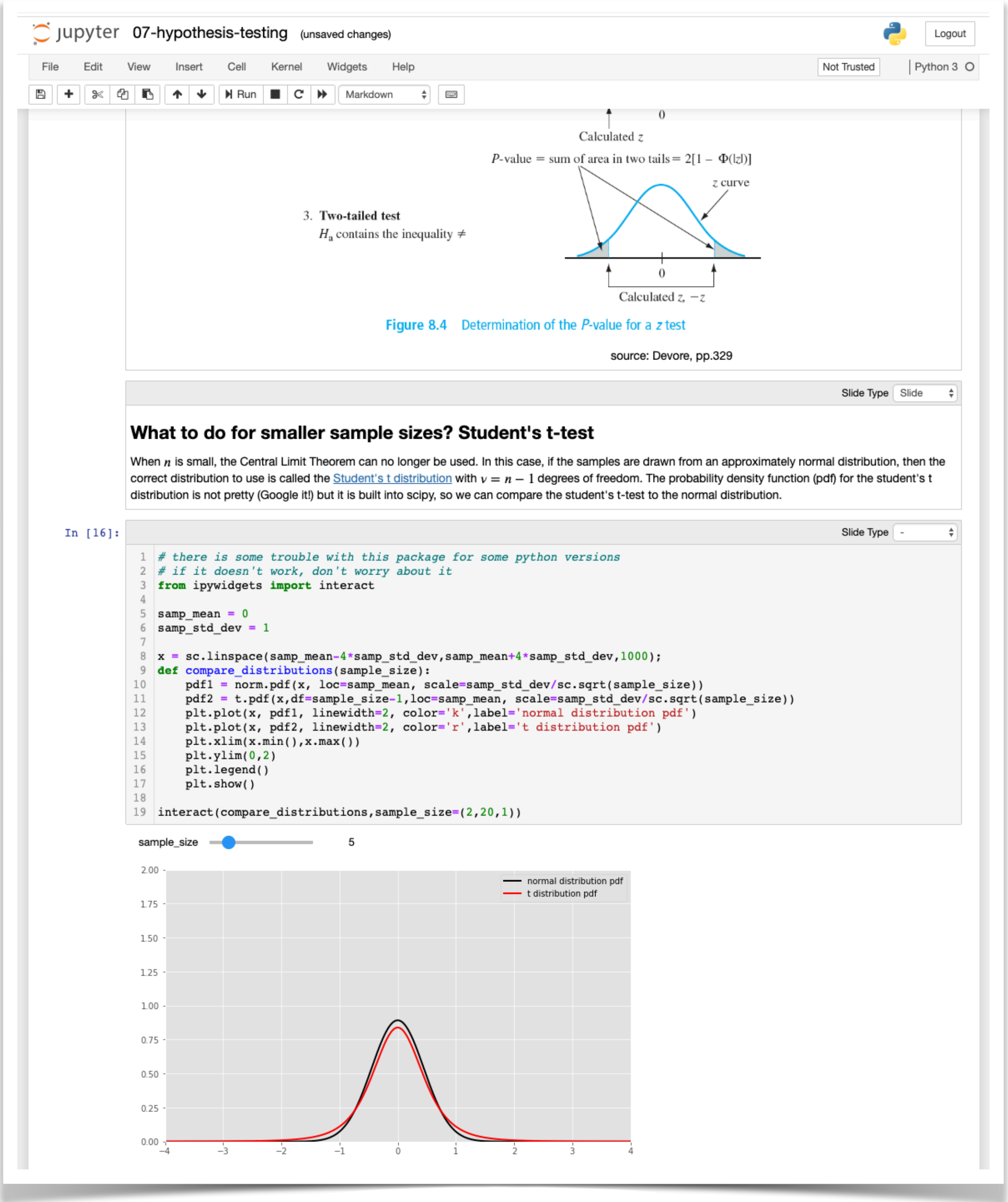
COMPUTATIONAL NOTEBOOKS: A MIDDLE GROUND?



Observable



R Markdown



Jupyter Notebooks

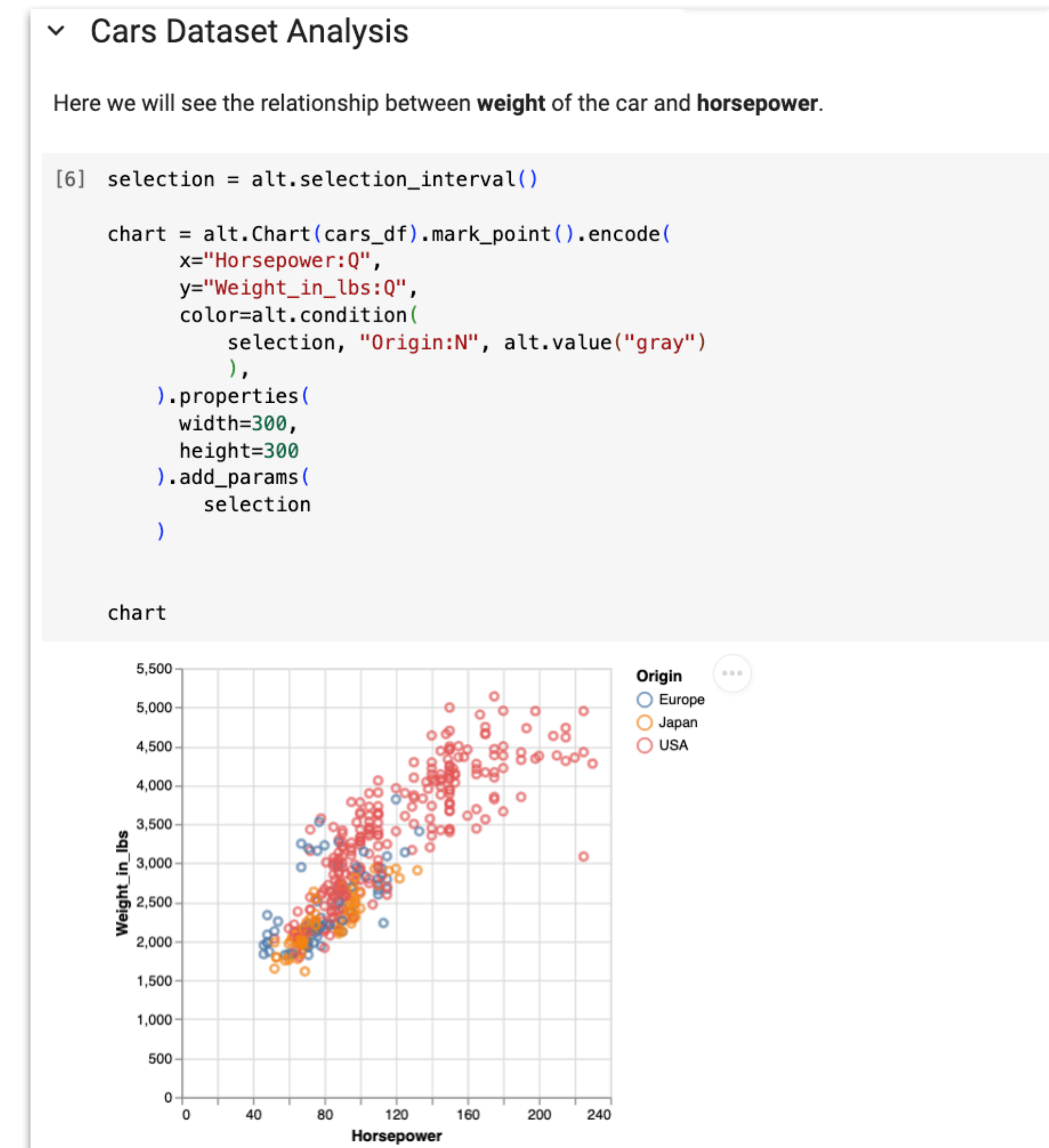
COMPUTATIONAL NOTEBOOKS: A MIDDLE GROUND?

Yes!

**Afford both scripting and
interactive visualization**

But visualizations are a dead end

can't "use" interaction in code
e.g., changing a label, or filtering
a value



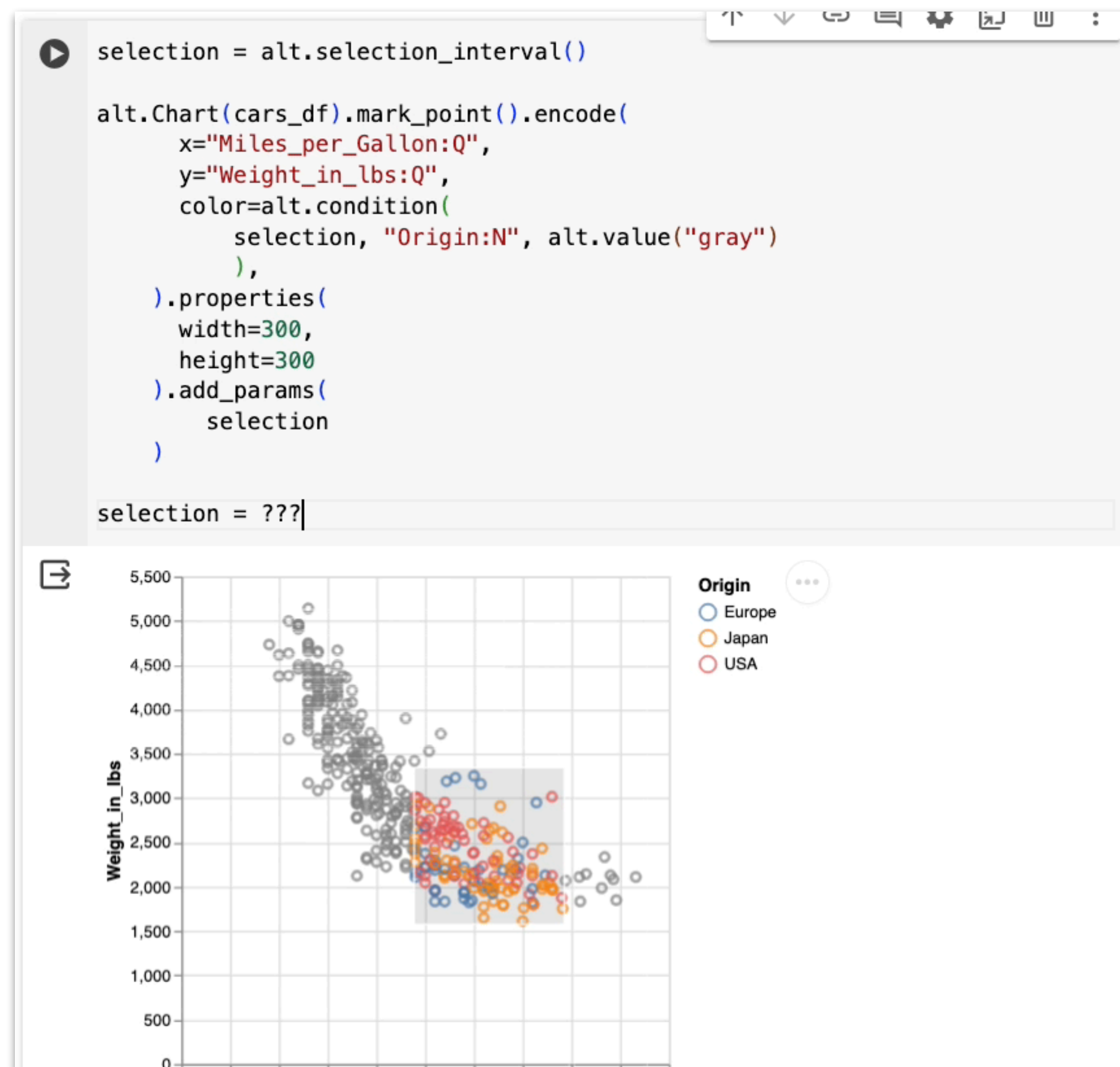
Documentation

Code

Visualizations

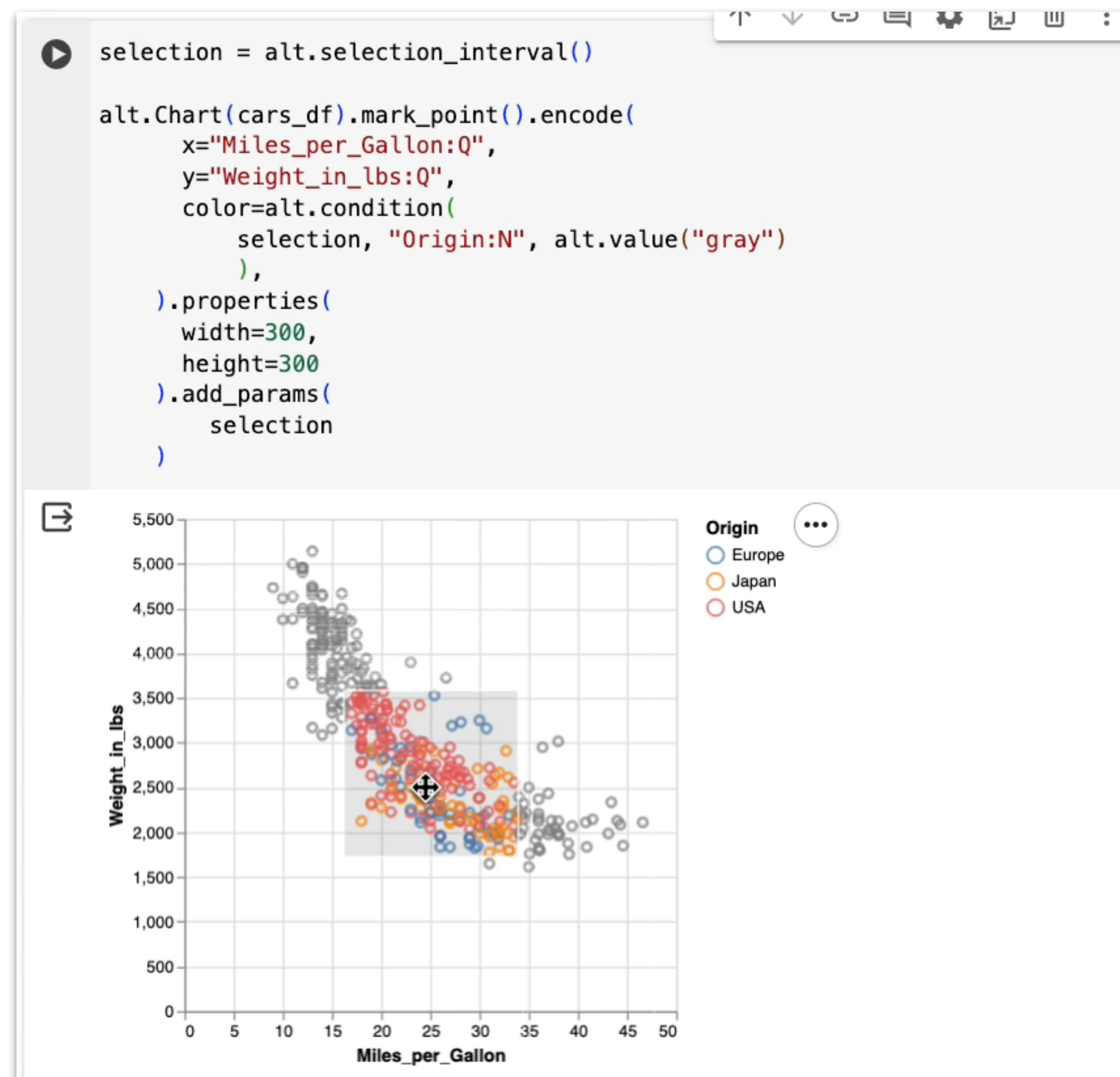
GAPS BETWEEN CODE AND INTERACTIVE OUTPUTS*

Semantic Gap



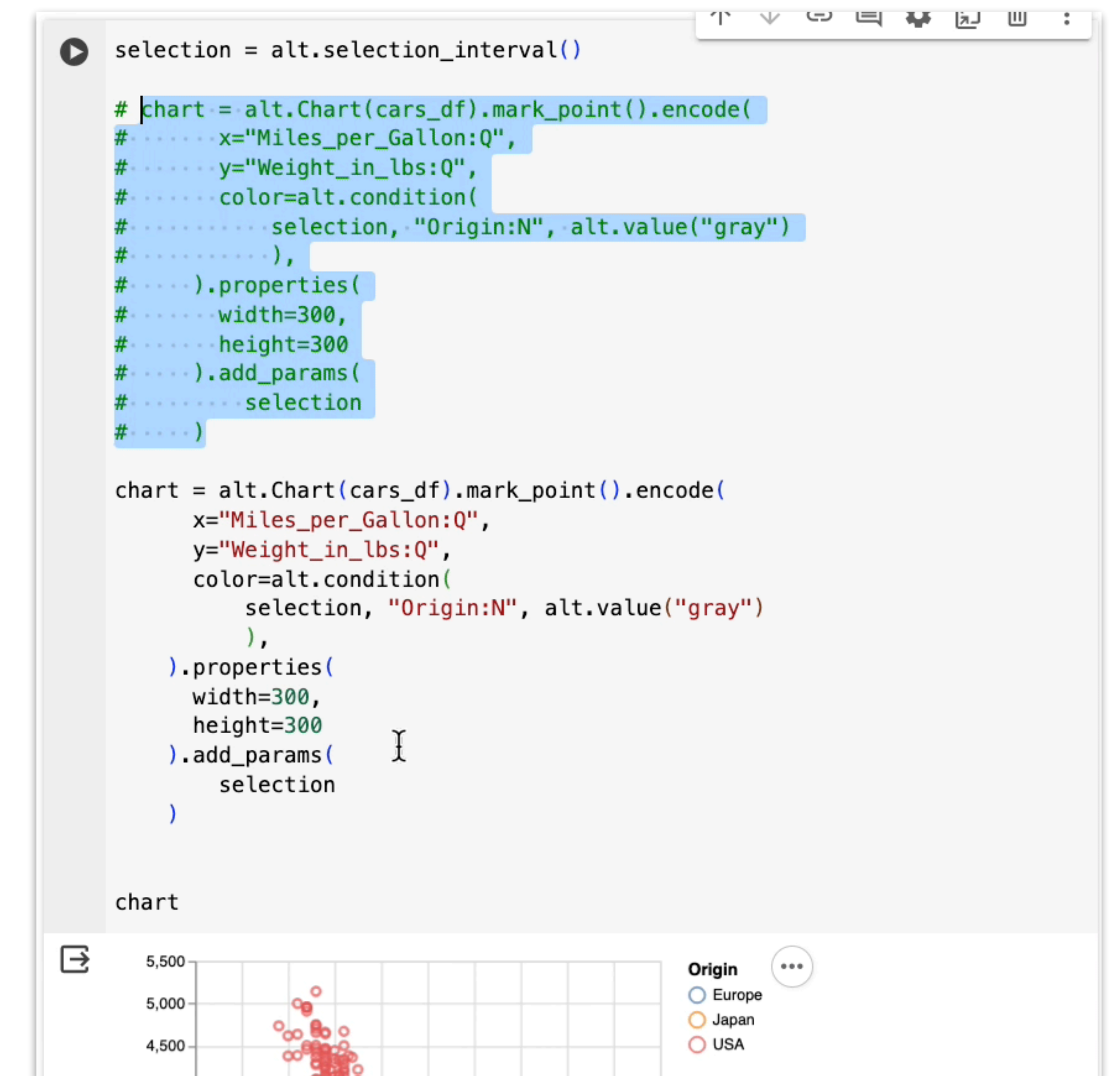
Information only flows from
code to visualization

Temporal Gap



Changes made to code are preserved
Changes made to vis are lost

Layout Gap



Changes in code are messy

*[Wu, Hellerstein, Satyanarayan, UIST 2020]

THESIS: **BRIDGING** BETWEEN CODE
AND INTERACTIVE VIS IS **USEFUL**

Easy handoffs are important!

THE **PERSIT** APPROACH

PRINCIPLE

Track events in interactive visualizations

Map them to **data frame operations**

Operations then **applied to data frame**

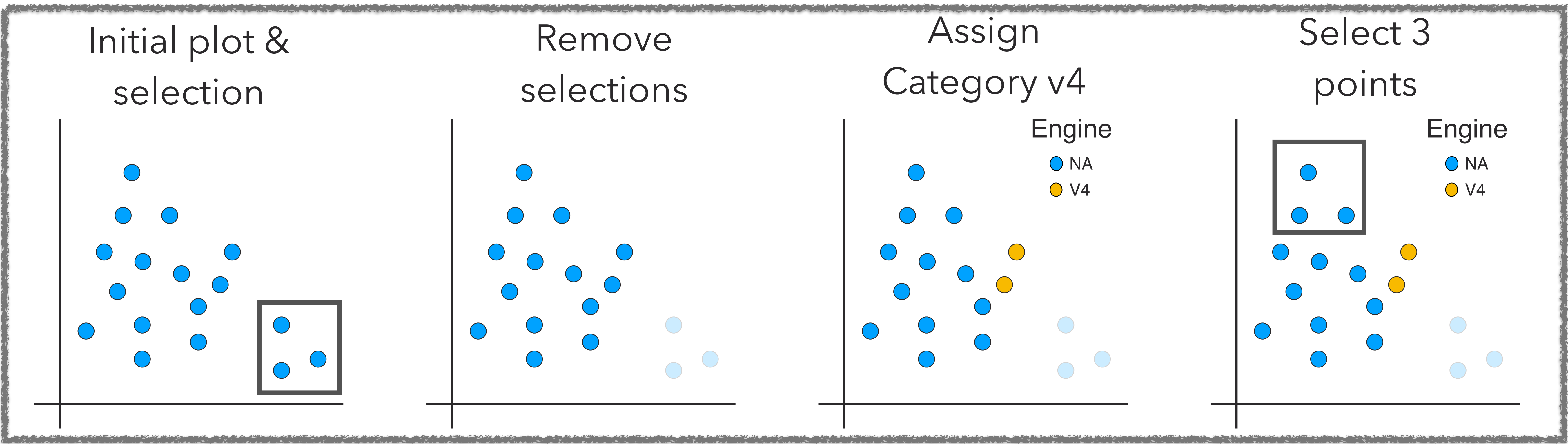
HOW IT WORKS

Code to create chart








```
df = pd.read_csv('cars.csv')  
PersistChart(scatterplot(df))
```

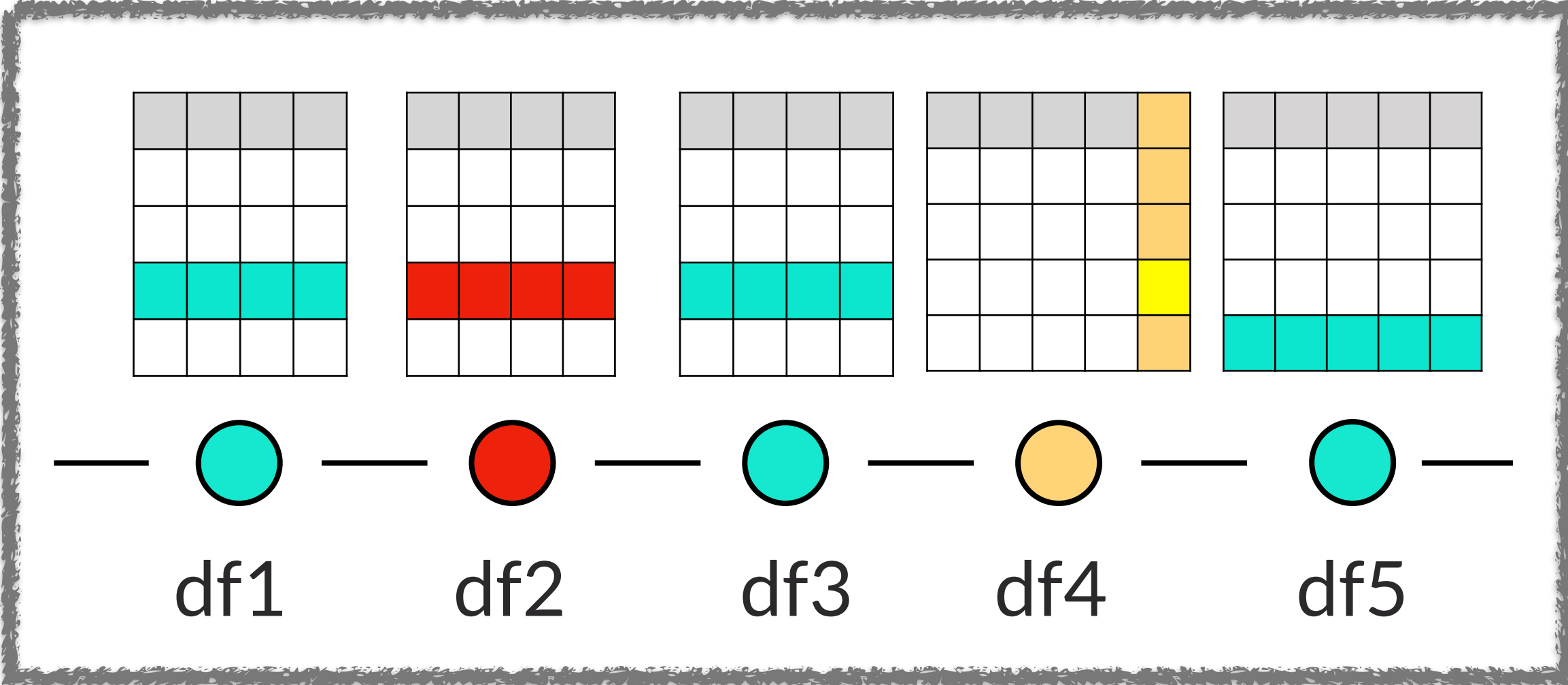
Interactive Visualization



Provenance

- df1  Select 3 points
- df2  Remove selections
- df3  Select 2 points
- df4  Assign category v4 to selection
- df5  Select 3 points

Dataframe updates



In code











```
df5.head()
```

Name	MPG	Cylinders	origin	Engine	is_selected
ford	18	8	USA	NA	FALSE
dodge	15	8	USA	NA	FALSE
volkswagen	22	4	Europe	V4	FALSE
amc	16	8	USA	NA	TRUE

27 rows x 8 columns

OPERATIONS

-  **Selection**
-  **Edit column names, edit cells**
-  **Sort rows/columns**
-  **Drop columns**
-  **Filter (in/out) items**
-  **Label items**
-  **Categorize items**
-  **Change data types**

PERSIST WORKFLOW

**Standard
Vega-Altair Chart**

```
import altair as alt
from vega_datasets import data
import persist_ext as PR

cars_df = data.cars() # Load cars dataset

brush = alt.selection_interval() # Create a 2d brush

# create scatterplot and link to brush
scatterplot = alt.Chart(cars_df).mark_point().encode(
    x="Miles_per_Gallon:Q",
    y="Weight_in_lbs:Q",
    color=alt.condition(brush, alt.value("steelblue"), alt.value("gray"))
).add_params(
    brush
)

PR.PersistChart(scatterplot)
```

Call to Persist with Chart as Parameter

```
import altair as alt
from vega_datasets import data
import persist_ext as PR

cars_df = data.cars() # Load cars dataset

brush = alt.selection_interval() # Create a 2d brush

# create scatterplot and link to brush
scatterplot = alt.Chart(cars_df).mark_point().encode(
    x="Miles_per_Gallon:Q",
    y="Weight_in_lbs:Q",
    color=alt.condition(brush, alt.value("steelblue"), alt.value("gray"))
).add_params(
    brush
)

PR.PersistChart(scatterplot)
```

Persist Toolbar

↶

↷

✎

✖

⊞

⊞

🔍

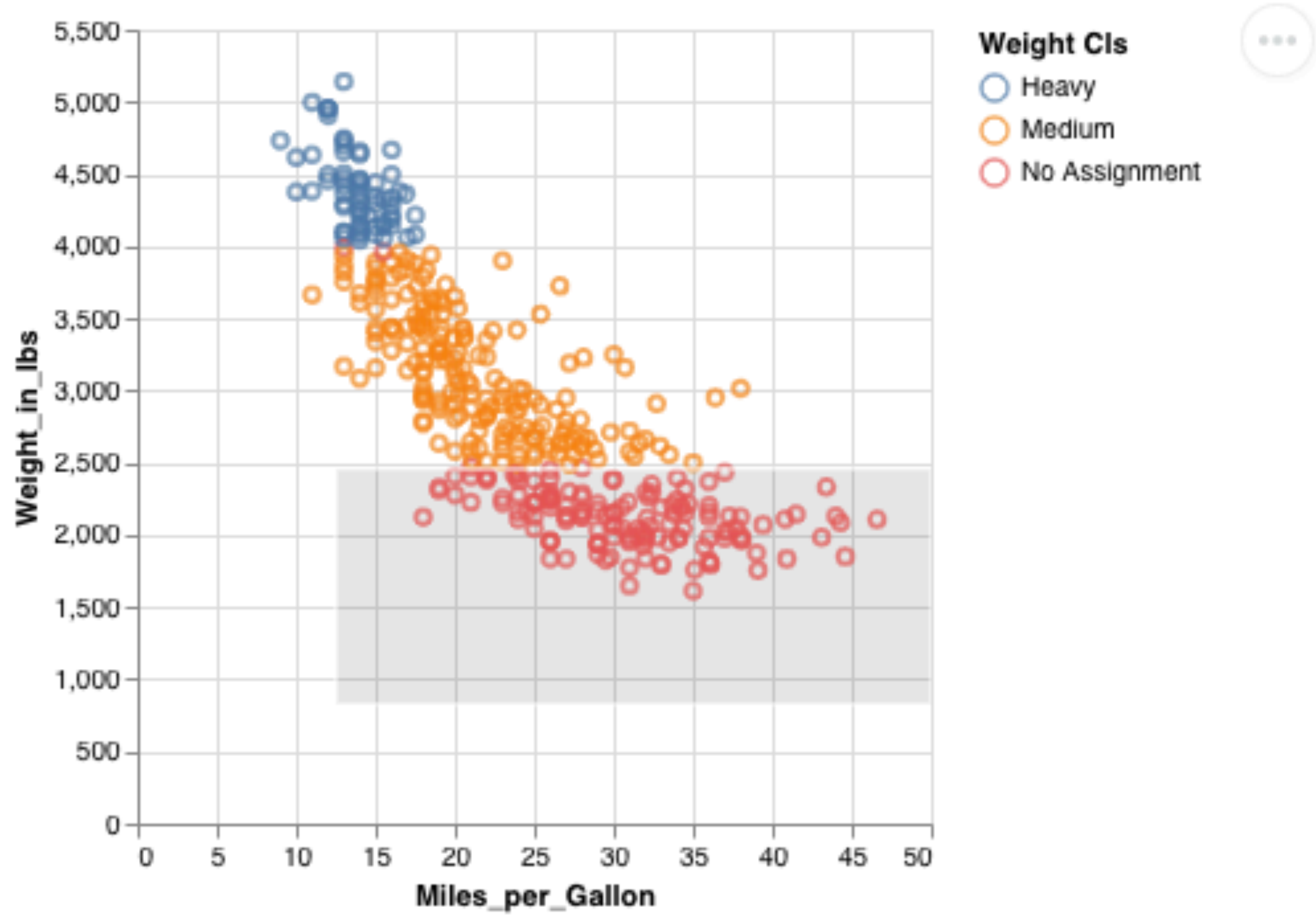
🔍

📄

Reset Trrack

Delete datasets

Vega-Altair Chart



Dataframe Manager

Dataframe name...

📄

+

🚀 persist_df_2

📄

+

Provenance History

Trrack

Summary

○ Root

○ Add new category 'We...

○ Add new option 'Heav...

○ Add new option 'Medi...

○ Add new option 'Light'...

○ Selected Miles_per_G...

○ Assign Weight Cls (He...

○ Selected Miles_per_G...

○ Assign Weight Cls (Me...

● Selected Miles_per_G...

○ Assign Weight Cls (Lig...

Dataframes:

🚀 persist_df_2

📄

+


```

import altair as alt
from vega_datasets import data
import persist_ext as PR

cars_df = data.cars() # Load cars dataset

brush = alt.selection_interval() # Create a 2d brush

# create scatterplot and link to brush
scatterplot = alt.Chart(cars_df).mark_point().encode(
    x="Miles_per_Gallon:Q",
    y="Weight_in_lbs:Q",
    color=alt.condition(brush, alt.value("steelblue"), alt.value("gray"))
).add_params(
    brush
)

PR.PersistChart(scatterplot)

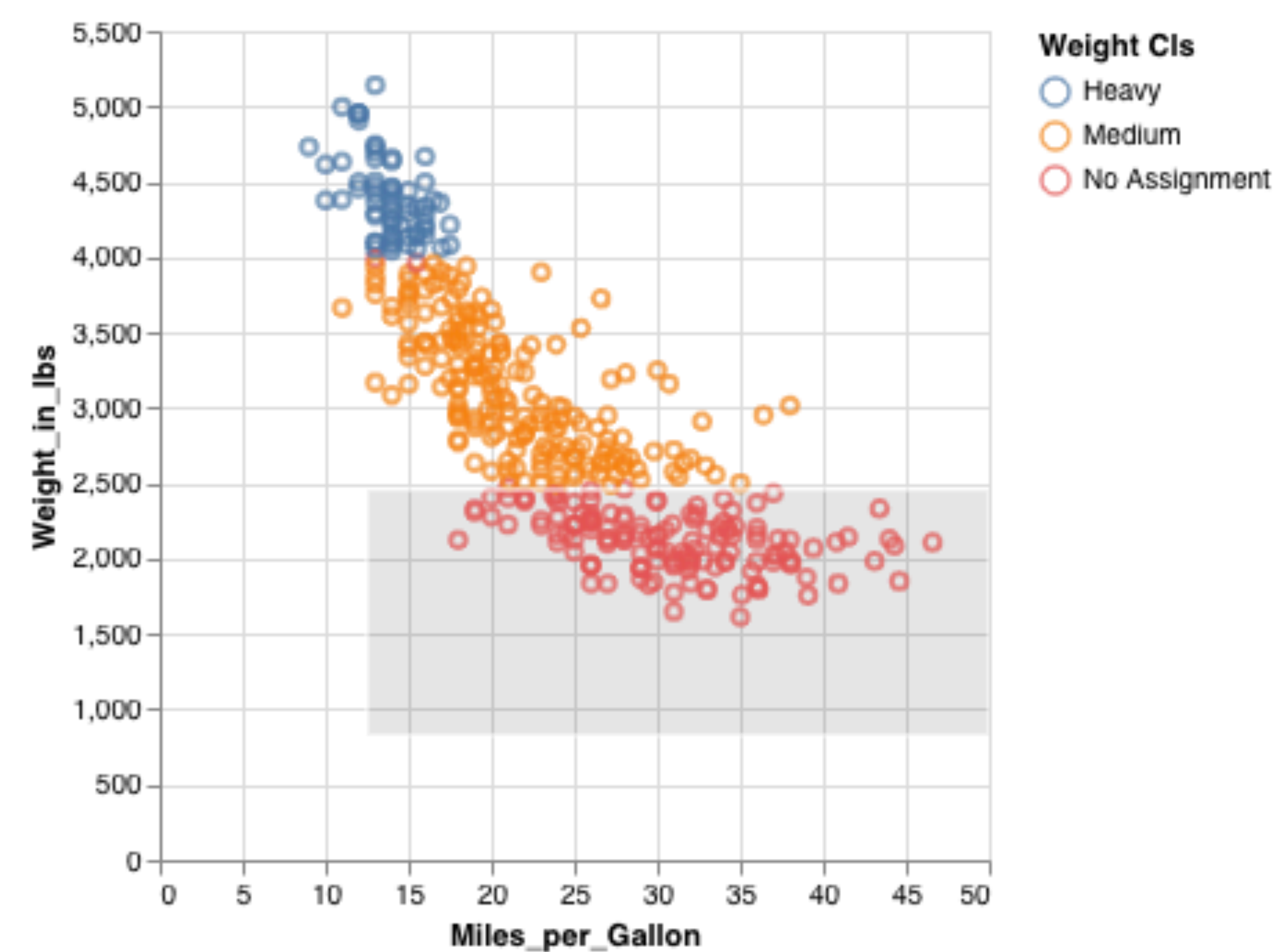
```

Categorization

Filtering

Labelling

[Reset Ttrack](#) [Delete datasets](#)



Dataframe name...

persist_df_2

Ttrack Summary

- Root
- Add new category 'We...
- Add new option 'Heav...
- Add new option 'Medi...
- Add new option 'Light'...
- Selected Miles_per_G...
- Assign Weight Cls (He...
- Selected Miles_per_G...
- Assign Weight Cls (Me...
- Selected Miles_per_G...

Dataframes:

persist_df_2

Assign Weight Cls (Lig...

Branches and **Choosing a State** in
provenance support *non-linear*
analysis, addressing **the layout gap**

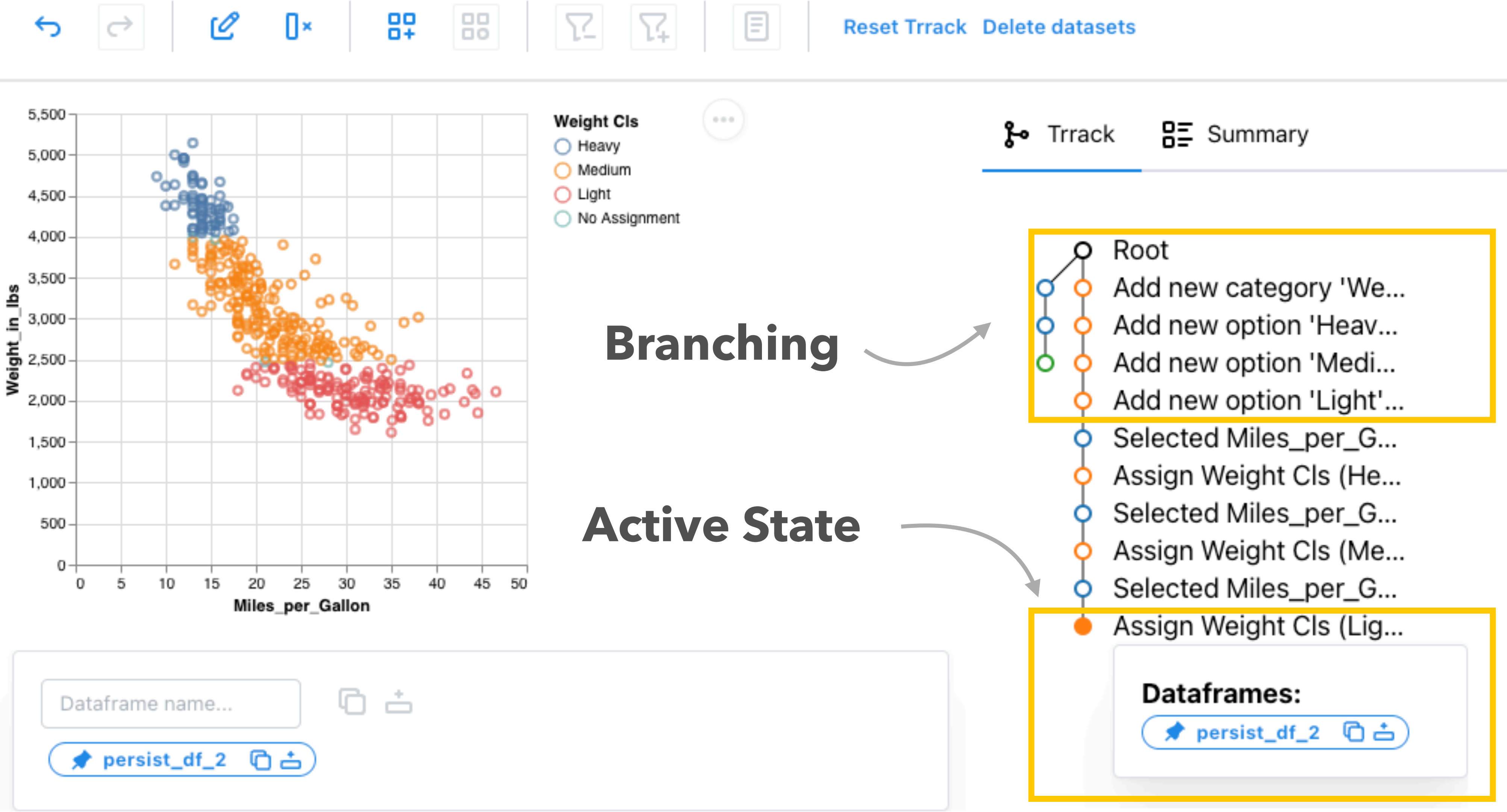
```
import altair as alt
from vega_datasets import data
import persist_ext as PR

cars_df = data.cars() # Load cars dataset

brush = alt.selection_interval() # Create a 2d brush

# create scatterplot and link to brush
scatterplot = alt.Chart(cars_df).mark_point().encode(
    x="Miles_per_Gallon:Q",
    y="Weight_in_lbs:Q",
    color=alt.condition(brush, alt.value("steelblue"), alt.value("gray"))
).add_params(
    brush
)

PR.PersistChart(scatterplot)
```



Branches and **Choosing a State** in
provenance support *non-linear*
analysis, addressing **the layout gap**

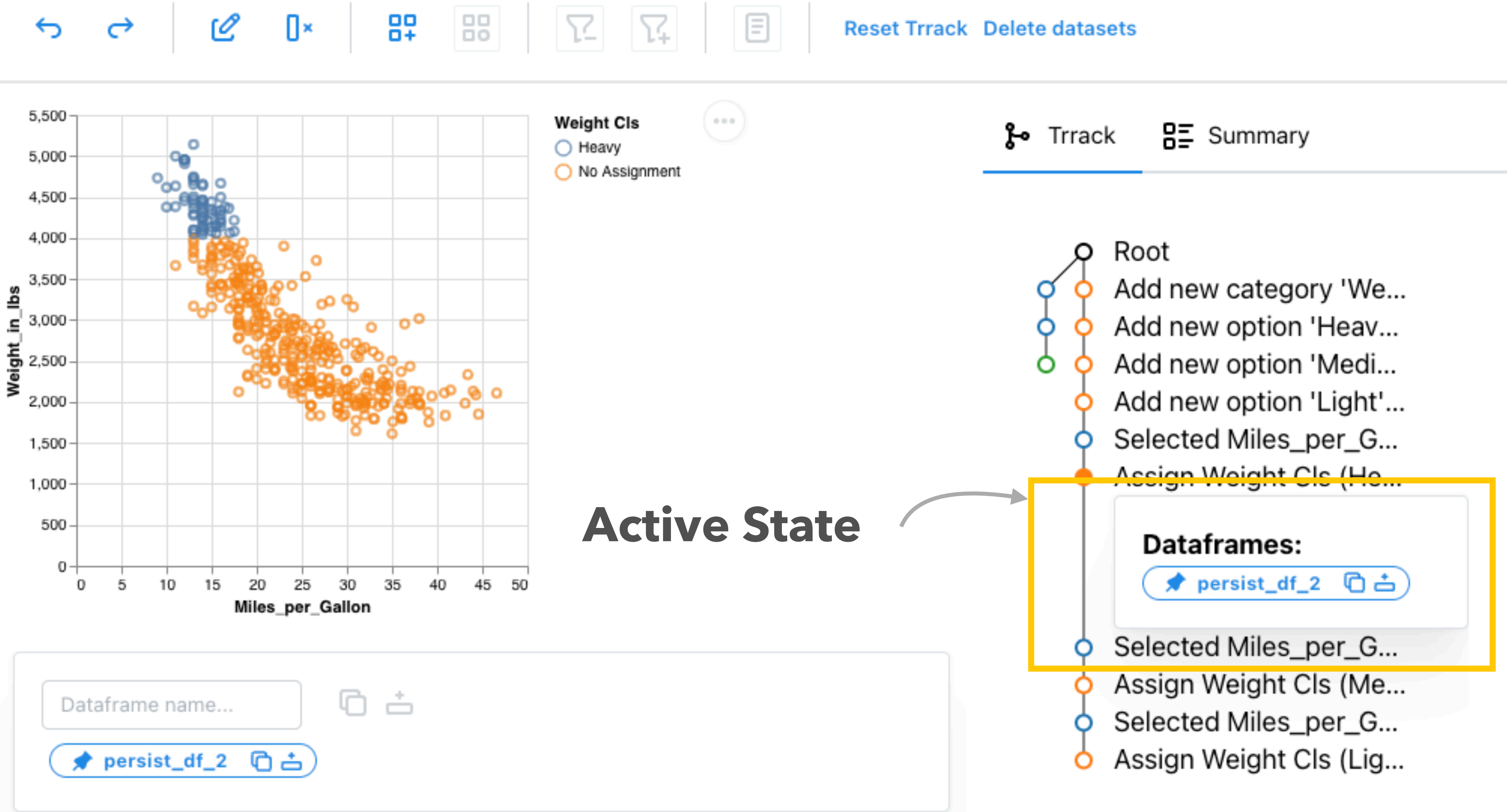
```
import altair as alt
from vega_datasets import data
import persist_ext as PR

cars_df = data.cars() # Load cars dataset

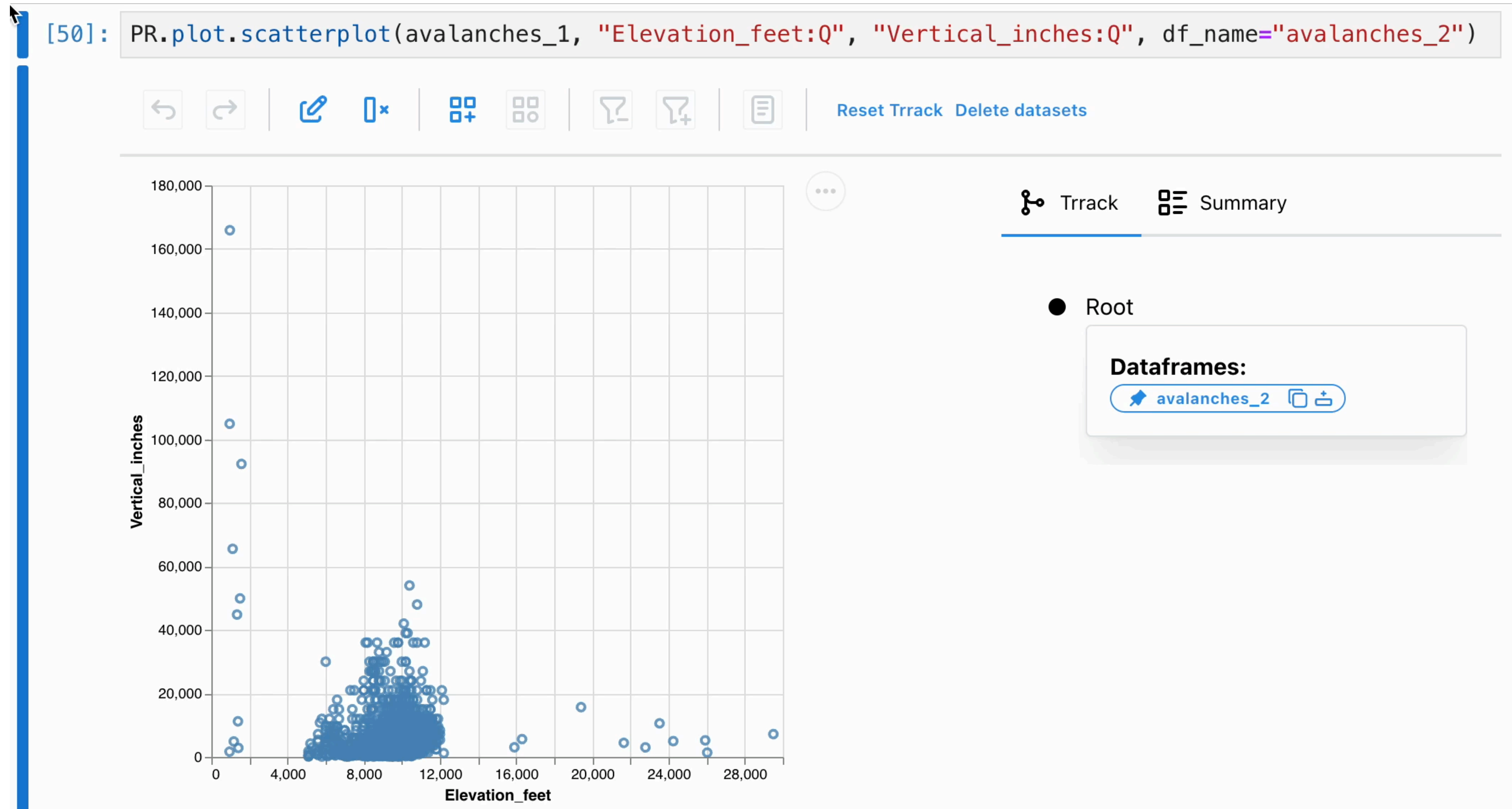
brush = alt.selection_interval() # Create a 2d brush

# create scatterplot and link to brush
scatterplot = alt.Chart(cars_df).mark_point().encode(
    x="Miles_per_Gallon:Q",
    y="Weight_in_lbs:Q",
    color=alt.condition(brush, alt.value("steelblue"), alt.value("gray"))
).add_params(
    brush
)

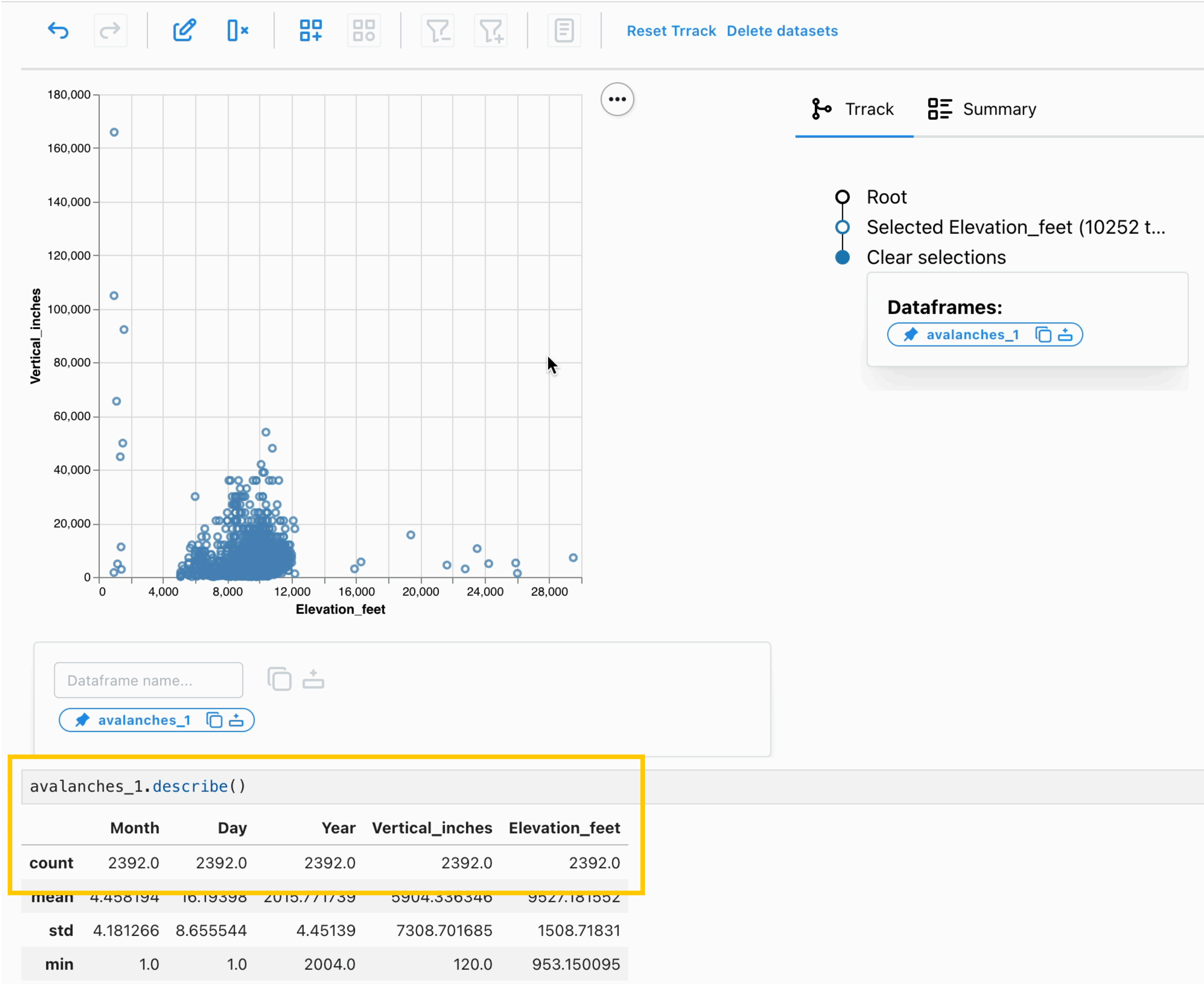
PR.PersistChart(scatterplot)
```



Persist **re-runs the interactions** in the output,
addressing **the temporal gap**



Persist *applies interactions to data frames* that can be accessed in code, addressing the **semantic gap**



**Manipulated
data frame accessed in
code**

count: 2392→79

VISUALIZATION OPTIONS



Arbitrary Vega-Altair Charts

The screenshot displays an interactive data table interface. The table has columns: #, Date, Place, Trigger, Weak Layer, Depth_inches, Aspect, and Elevation_feet. The first three rows are visible: 2338 (3-23-2023, Dry Creek, Natural), 955 (1-19-2014, Whitney Basin, Snowmobiler), and 1028 (2-21-2014, Chalk Creek, Natural). A 'Track' panel on the right shows a sequence of operations: Root, Rename column ;Weak Lay..., Drop column Comment..., Updated column 'Depth_i...', Updated column 'Depth_i...', Updated column 'Depth_i...', Changed column 'Dept...', Sort (descending) by 'Dep...', Drop column Coordinates, Selected 1 point, Selected 2 points, Selected 3 points, Drop column ;Region, Drop column Width_inches, and Drop column Vertical_inch... A 'Dataframes' panel at the bottom shows 'current_df' as the current dataframe.

#	Date	Place	Trigger	Weak Layer	Depth_inches	Aspect	Elevation_feet
2338	3-23-2023	Dry Creek	Natural			West	8000
955	1-19-2014	Whitney Basin	Snowmobiler			East	10500
1028	2-21-2014	Chalk Creek	Natural			Northeast	10600
1024	2-17-2014	Upper Weber Canyon	Natural			Northeast	10400
998	2-12-2014	Upper Weber Canyon	Natural			Northeast	10400
938	1-14-2014	Upper Weber Canyon	Explosive			East	10300
1299	1-26-2016	Currant Creek Peak	Snowmobiler			Southwest	9500
1044	2-28-2014	Chalk Creek	Natural			Northeast	10600
2348	3-30-2023	Bunnels	Natural			Northeast	10800
1977	4-6-2021	Blue Ice	Natural			Northeast	10400

An Interactive Data Table

VEGA-ALTair

Persist works with **most Vega-Altair** charts

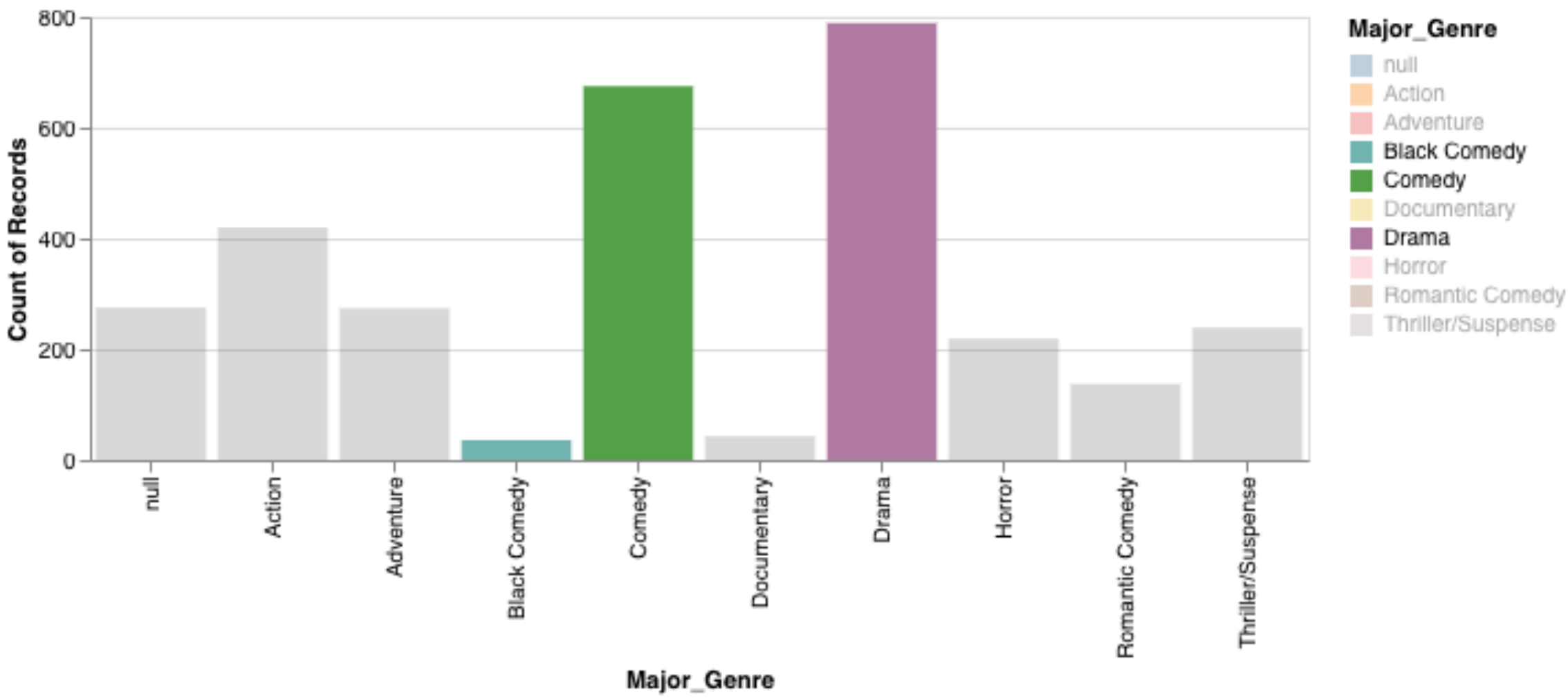
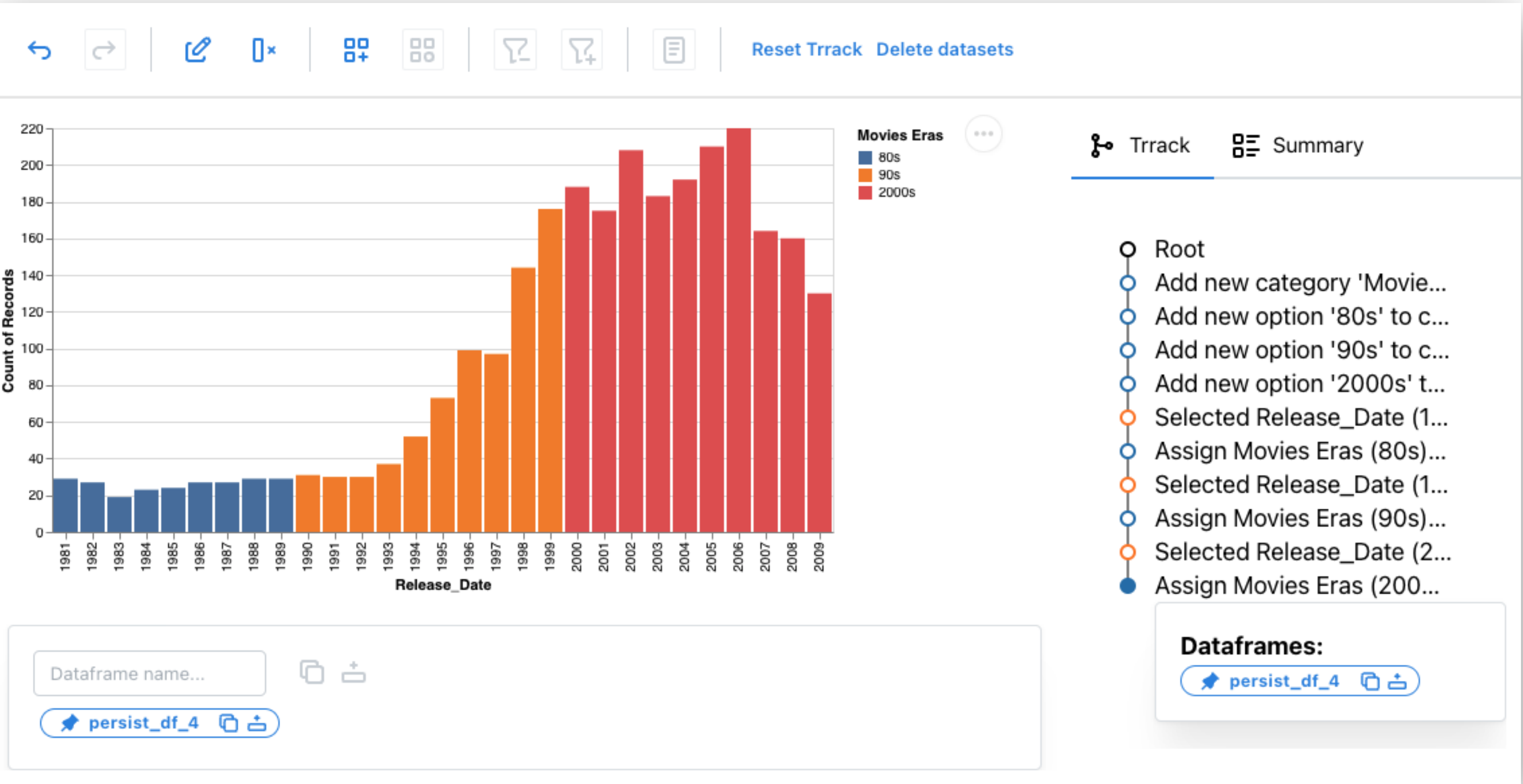
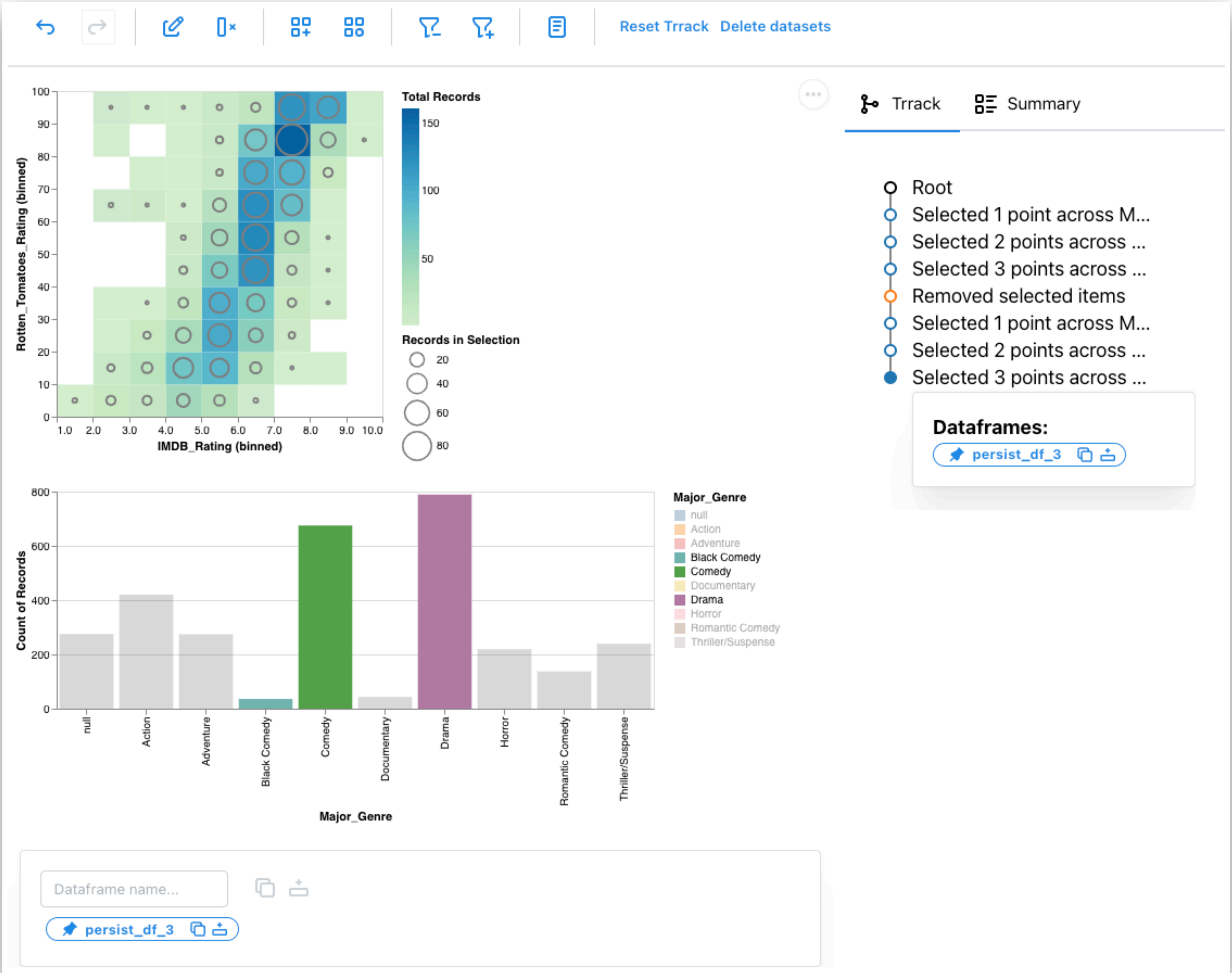
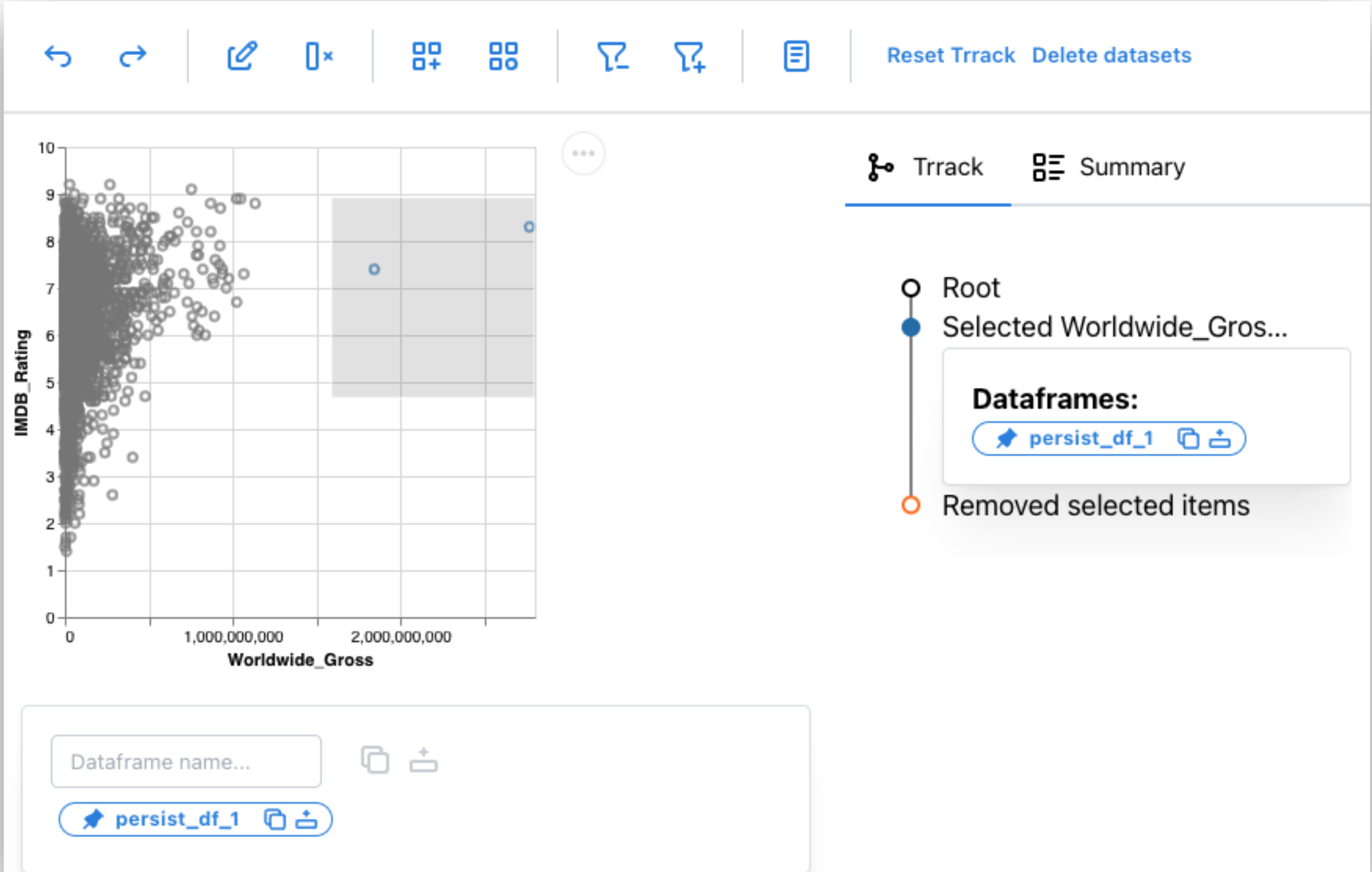
"Listens" to native operations (selections)

Updates Vega charts:

- Use original chart spec when possible (e.g., filters)

- Update spec when necessary (categories, labels)

EXAMPLE CHARTS



DEMO

EVALUATION

IN-LAB STUDY

III STUDY DESIGN

WE RECRUITED ELEVEN PARTICIPANTS FOR THE STUDY. PARTICIPANTS ALL HAD PRIOR EXPERIENCE WITH PYTHON AND PANDAS.

FULL FACTORIAL DESIGN

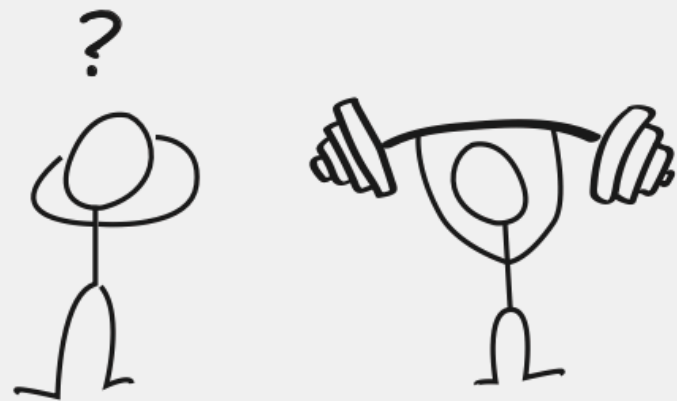
2 DATASETS **X** 2 CONDITIONS **X**

THE ORDER OF CONDITIONS WAS RANDOMLY ASSIGNED.

FOR EACH CONDITION, DATASETS WERE RANDOMLY ASSIGNED. PARTICIPANTS NEVER SAW THE SAME DATASET TWICE



— STUDY METRICS —



SUBJECTIVE PERFORMANCE



TIME



ERROR



REPRODUCIBILITY

— CONDITIONS —

PANDAS CODING

```
[ 1 ] import pandas as pd
      df = pd.read_csv(...)
```

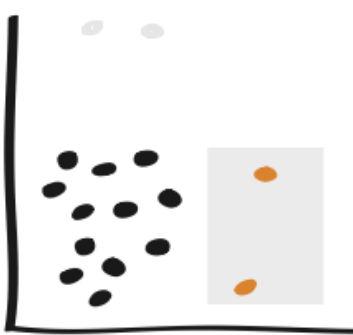
```
[ 2 ] df = df.drop("AGE")

      df = df.rename(
          columns={'JOB': 'PROFESSION'}
      )

      df.loc[1, 'JOB'] = 'PRINCIPAL'
```

PERSIST EXTENSION

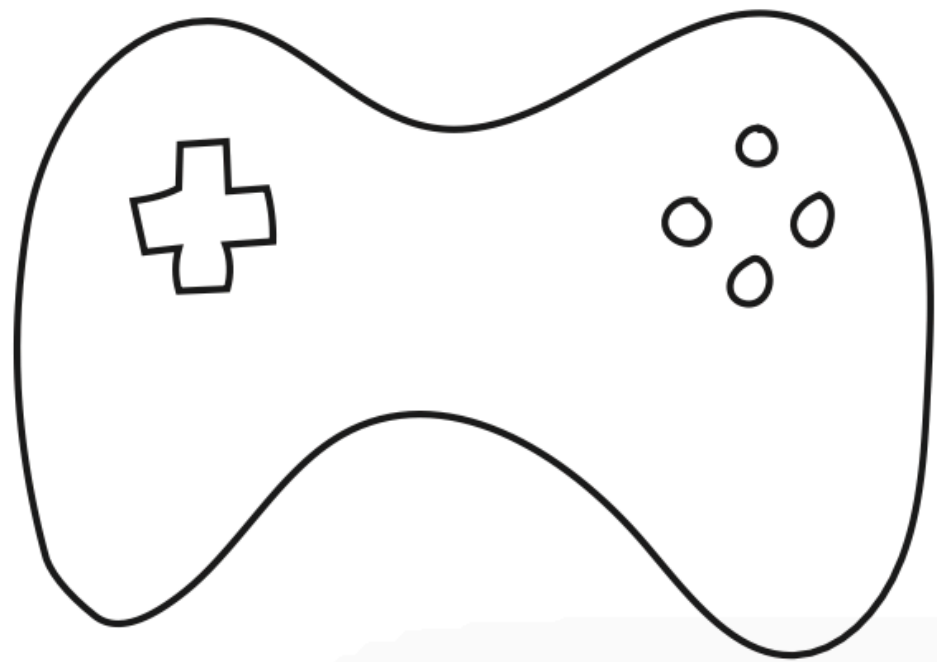
```
[ 1 ] import persist_ext as pr
      import pandas as pd
      df = pd.read_csv(...)
      pr.plot.scatterplot(...)
```



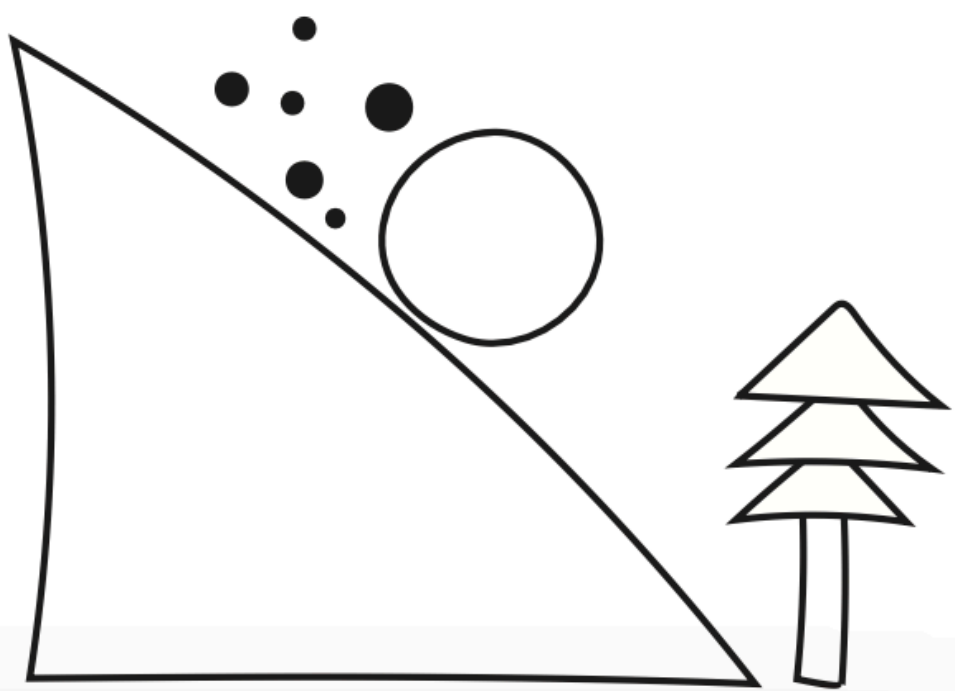
SELECT 2 VALUES
REMOVE 2 VALUES
SELECT 2 VALUES

— DATASETS —

VIDEO GAMES



AVALANCHES



IN-LAB STUDY

TASKS

PARTICIPANTS MADE THE FOLLOWING CHANGES TO A DATASET

NAME	AGE	JOB
STEVE	32	PLUMBER
JILL	24	TEACHER
ANN	42	ENGINEER

REMOVE COLUMNS

NAME	AGE	PROFESSION
STEVE	32	PLUMBER
JILL	24	TEACHER
ANN	42	ENGINEER

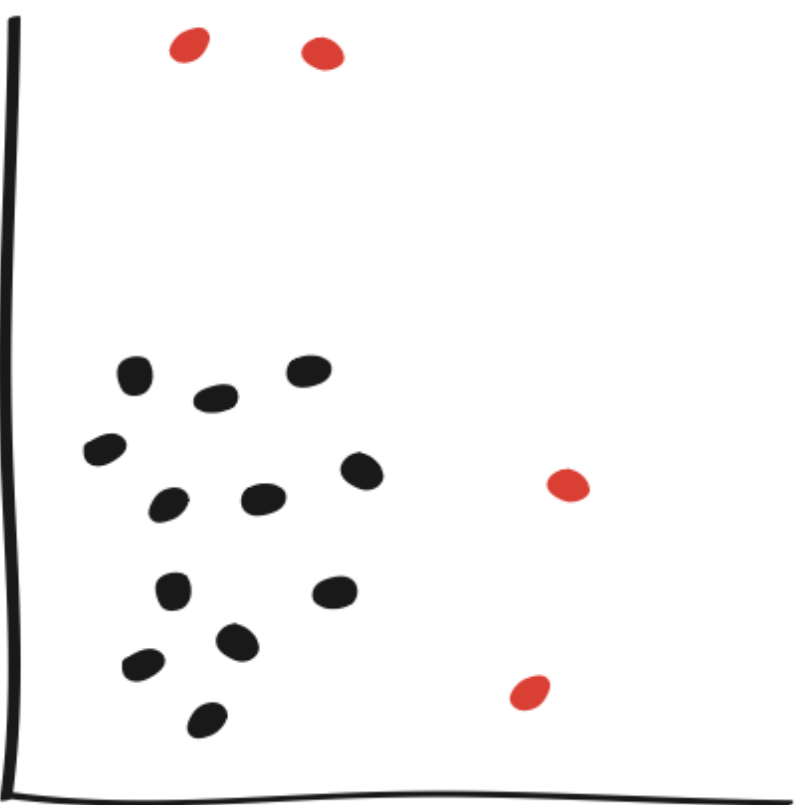
RENAME COLUMNS

AGE	AGE
"32"	32
"24"	24
"42"	42

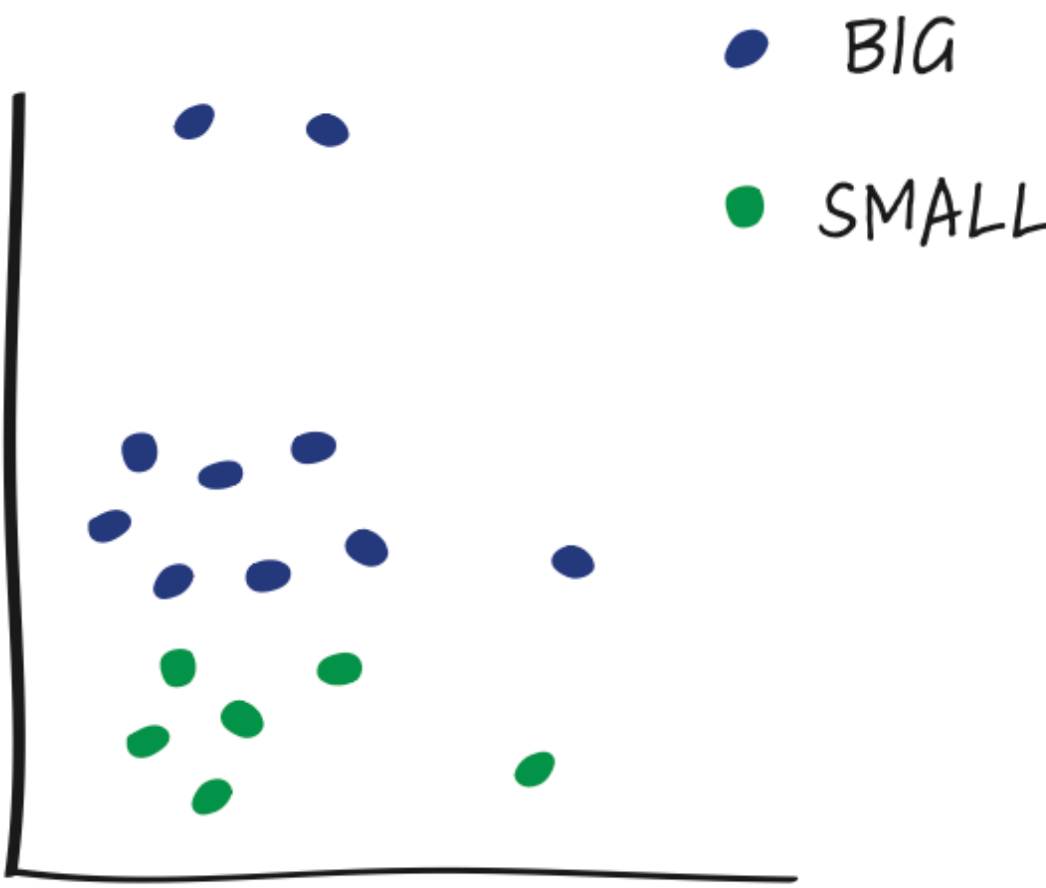
CHANGE DATA TYPE

NAME	AGE	JOB
STEVE	32	PLUMBER
JILL	24	PRINCIPAL
ANN	42	ENGINEER

EDIT VALUES



FILTER DATA



ADD CATEGORICAL
COLUMN

RESULTS

3x

times faster
with Persist

97%

tasks **correctly** using Persist,
compared to 85% for Pandas

11/11

notebooks using Persist
were **reproducible**

only 7/11 using pandas
were

RESULTS

QUOTES

“so much easier than manually coding.”
- M4

“easier as compared to the code and everything was visible [...] and it didn’t take much time.”
- M2

“Changing the category type, or adding new categories or removing anomalies from data, they were very much easier in [Persist] than coding.”
- M7

“The thing I really liked about is version control, which shows the history of all operations [...] and also saves the changes [...] into a data frame.”
- M14

DISCUSSION

GENERATING CODE VS PROVENANCE TRACKING

Provenance **better for most cases**

No code clutter

Undo/redo

Consistent semantics

But code generation might be

more robust

works outside of notebooks

works w/o the library

GENERATE CODE ON DEMAND!

[Beta]

↶

↷

✎

✖

⛶

⛶

📄

Reset Ttrack

Delete datasets

🔍 Search

☰

<input type="checkbox"/>	#	Region ↑↓ ⋮ ⋮	Year ↑↓ ⋮ ⋮	;Trigger ↑↓ ⋮ ⋮	;Weak Layer ↑↓ ⋮ ⋮	Dep
<input type="checkbox"/>	3	Salt Lake	2012	Skier		Facets
<input type="checkbox"/>	4	Salt Lake	2012	Skier		New Snow
<input type="checkbox"/>	5	Salt Lake	2012	Skier		Facets
<input type="checkbox"/>	6	Salt Lake	2012	Skier	New Snow/Old Snow Interface	
<input type="checkbox"/>	7	Salt Lake	2012	Skier		Facets
<input type="checkbox"/>	8	Salt Lake	2012	Skier		Facets
<input type="checkbox"/>	9	Salt Lake	2012	Skier		Facets
<input type="checkbox"/>	10	Salt Lake	2012	Skier		Facets
<input type="checkbox"/>	11	Salt Lake	2012	Skier		Facets
<input type="checkbox"/>	12	Salt Lake	2012	Unknown		Ground Interface

Rows per page

10

1-10 of 2,390

⏪

⏴

⏵

⏩

Dataframe name...

📄

+

🔍 persist_df_1

📄

+

```

def create_persist_df_1(df):
    df = df.copy(deep=True)

    # Add "__id_column" as the ID column
    df.insert(0, "__id_column", df.index + 1)
    df["__id_column"] = df["__id_column"].astype(str)

    # Add selection column
    df["__selected"] = False

    # Rename column
    df = df.rename(columns={';Region': 'Region'})
    df.loc[df["__id_column"].isin(['2']), "__selected"] = True
    df.loc[df["__id_column"].isin(['1', '2']), "__selected"] = True

```

📄 Copy

📄 Insert

🔑 Ttrack

☰ Summary

○ Root

○ Rename column ;Region to Re...

○ Selected 1 point

○ Selected 2 points

○ Removed selected items

Dataframes:

🔍 persist_df_1

📄

+

TRY OUT PERSIST!

Persist is **available now!**

<https://vdl.sci.utah.edu/persist/>

Documentation & examples

Feedback / bug reports
appreciated!

README BSD-3-Clause license

Getting Started

Requirements

- JupyterLab >= 4.0.0 or Jupyter Notebook >= 7.0.0
- pandas >= 0.25
- altair >= 5
- ipywidgets
- anywidget

Install

To install the extension, execute:

```
pip install persist_ext
```

If the Jupyter server was already running, you might have to reload the browser page and restart the kernel.

Uninstall

To remove the extension, execute:

```
pip uninstall persist_ext
```

Usage

Persist supports two types of interactive outputs — a custom data table and [Vega-Altair](#) (>=5.0.0, see [requirements](#) and [caveats](#)) charts. The following examples will walk you through creating each one. The examples are also available as notebooks in the `examples` folder of the repository. Each section will link to the corresponding notebook as well as a binder link for the notebook.

Persist currently works with pandas dataframes, so load/convert the data to pandas dataframe before using.

Examples

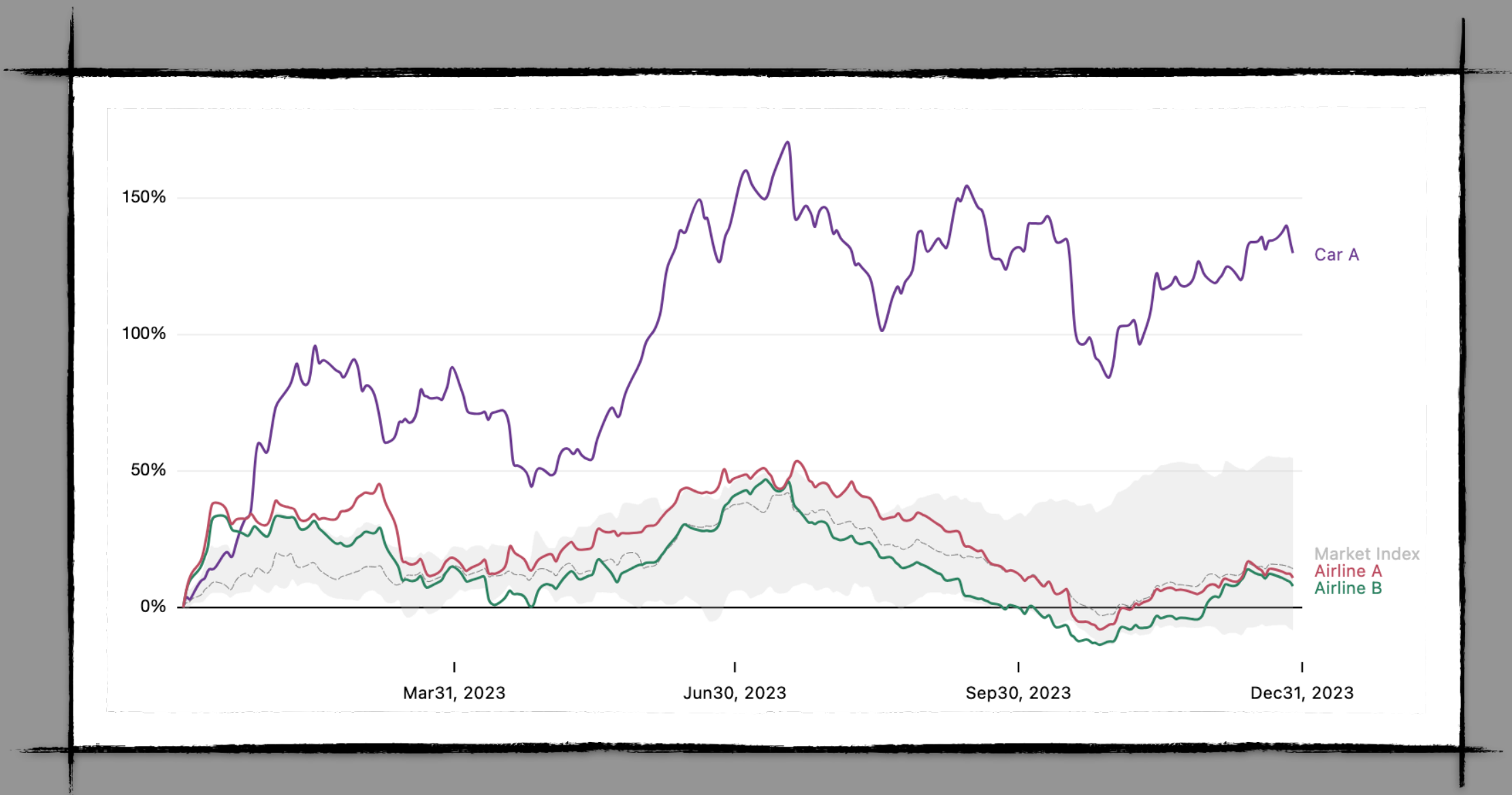
EMPIRICAL & THEORETICAL WORK



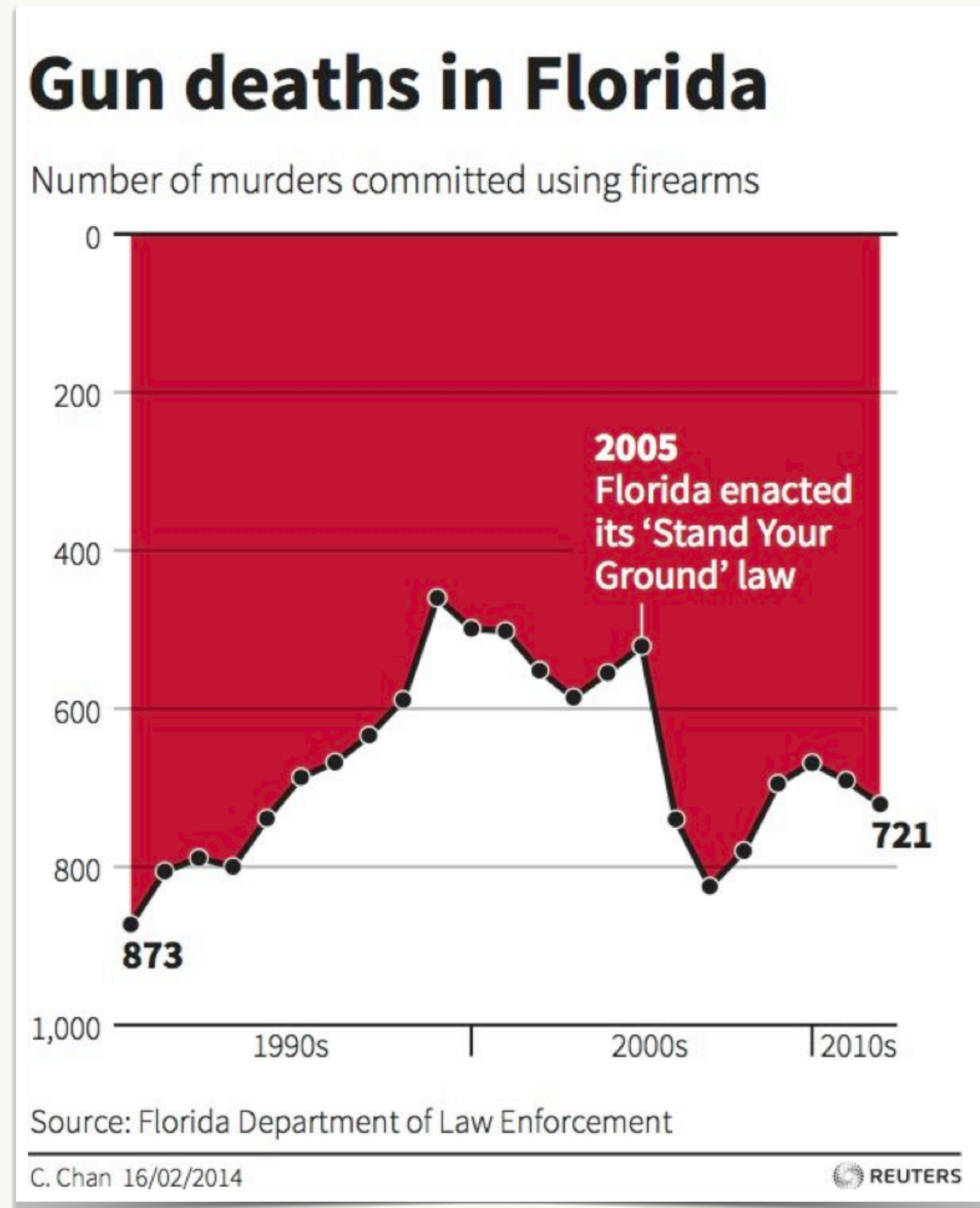
Max Lisnic, Marina Kogan

HOW PEOPLE ACTUALLY LIE WITH CHARTS...

... AND WHAT WE CAN DO ABOUT IT



HOW WE THINK PPL LIE WITH CHARTS: VIOLATIONS OF VISUALIZATION DESIGN GUIDELINES



Inverted y axis

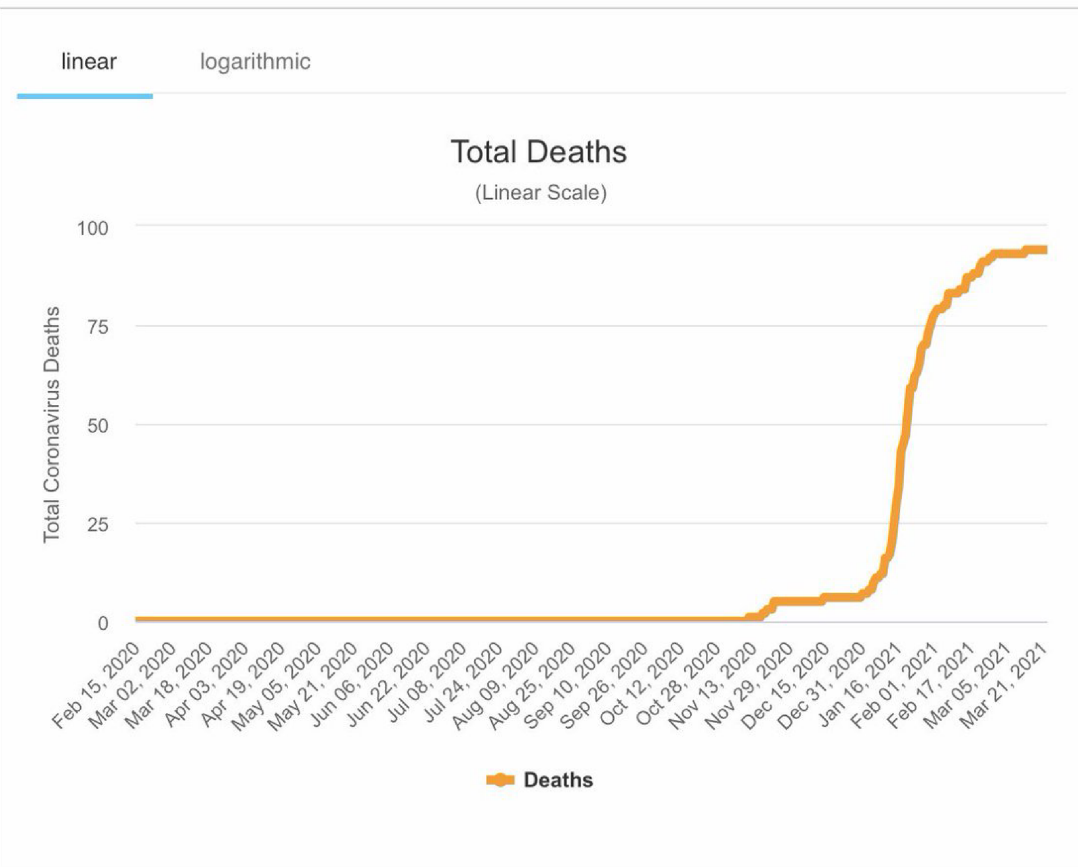


Truncated y axis

**BUT IS THAT REALLY HOW
PEOPLE LIE WITH CHARS?**

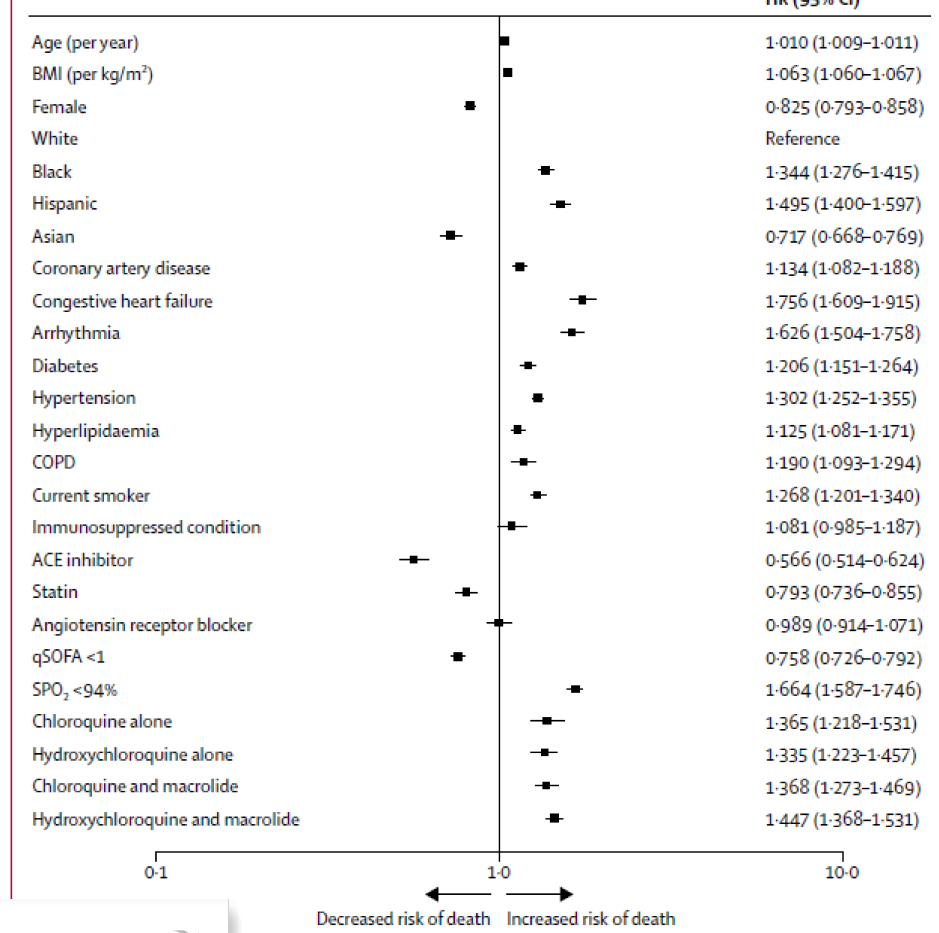
As of 10th January when they started vaccinations there had only been 16 cases.

Total Coronavirus Deaths in Gibraltar

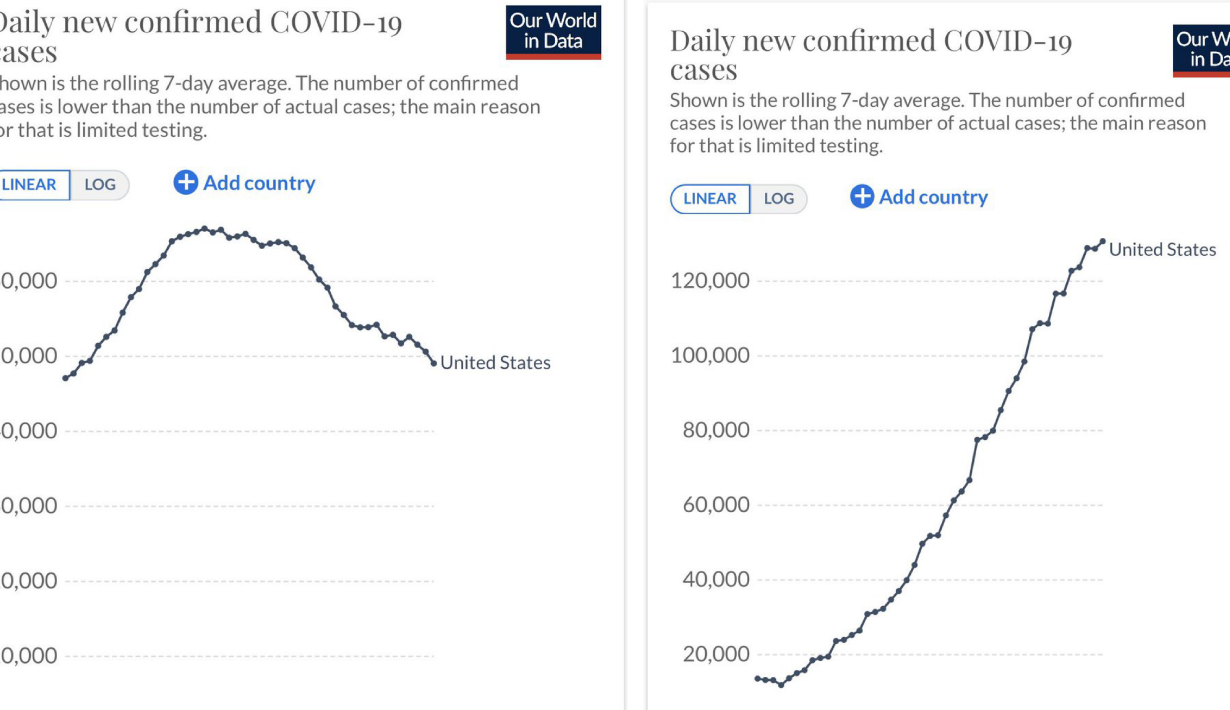


Taking HCQ is as strongly associated with increased coronavirus death risk as DIABETES.

[thelancet.com/journals/lance...](https://www.thelancet.com/journals/lance...)

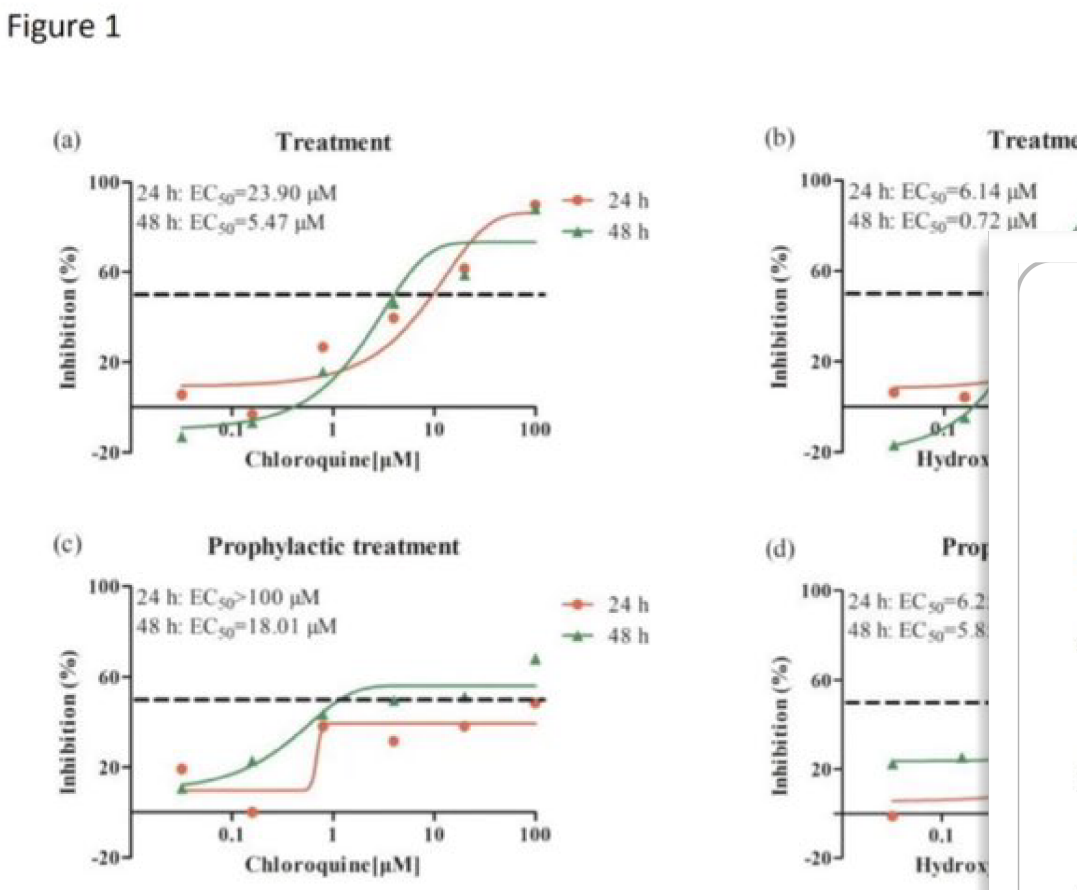


New daily covid cases this time last year (pre-vaccination) vs this year (post-vaccination)



Hydroxychloroquine probably better

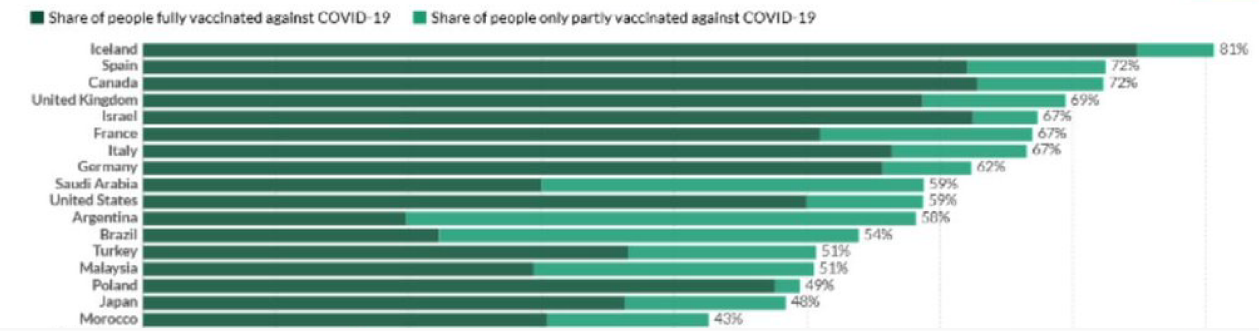
m.youtube.com/watch?feature=...



Most vaccinated: Iceland 81%
Least vaccinated: Nigeria 1.2%

Iceland has 119 times more Covid cases

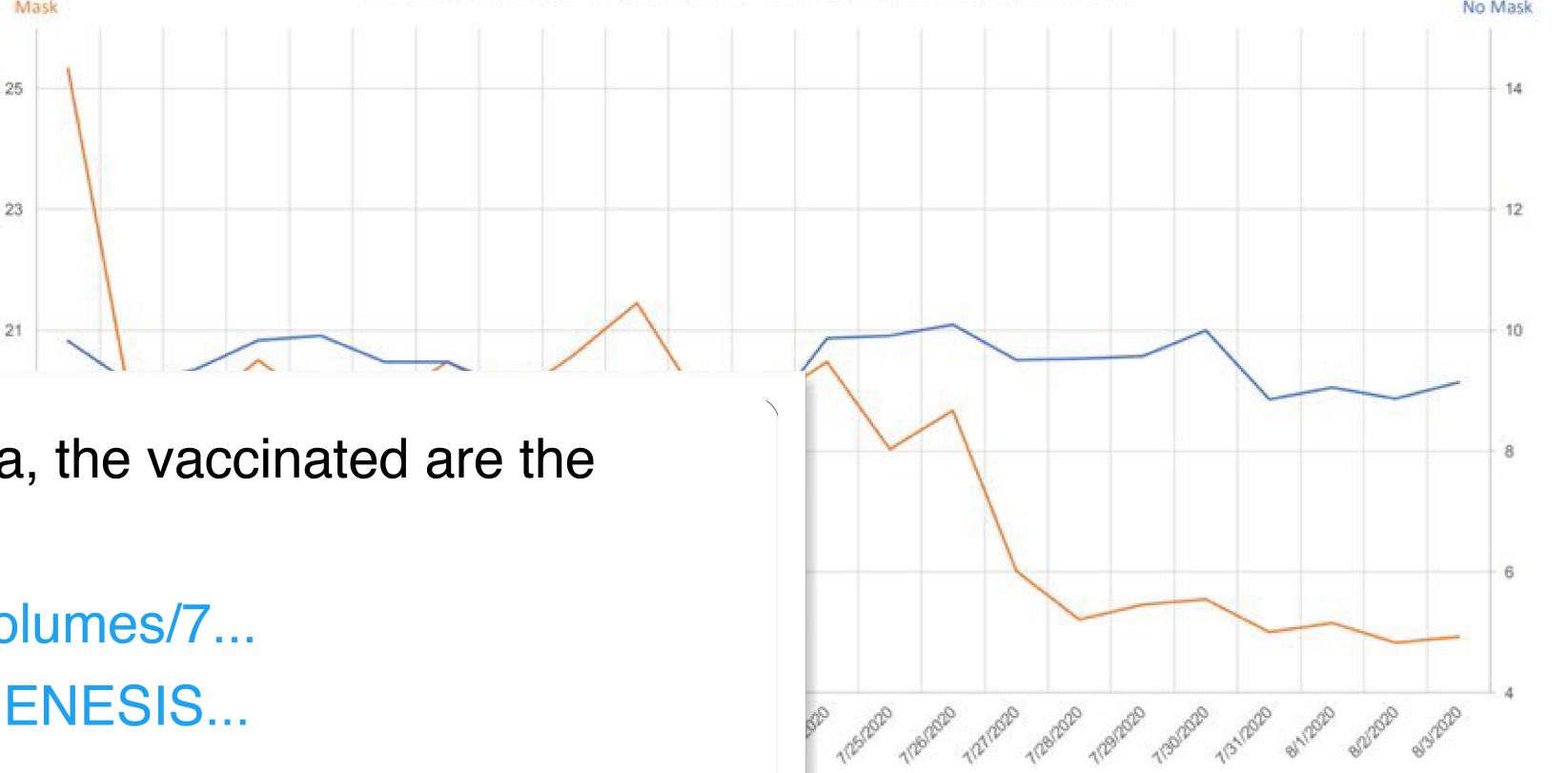
Share of people vaccinated against COVID-19, Aug 11, 2021



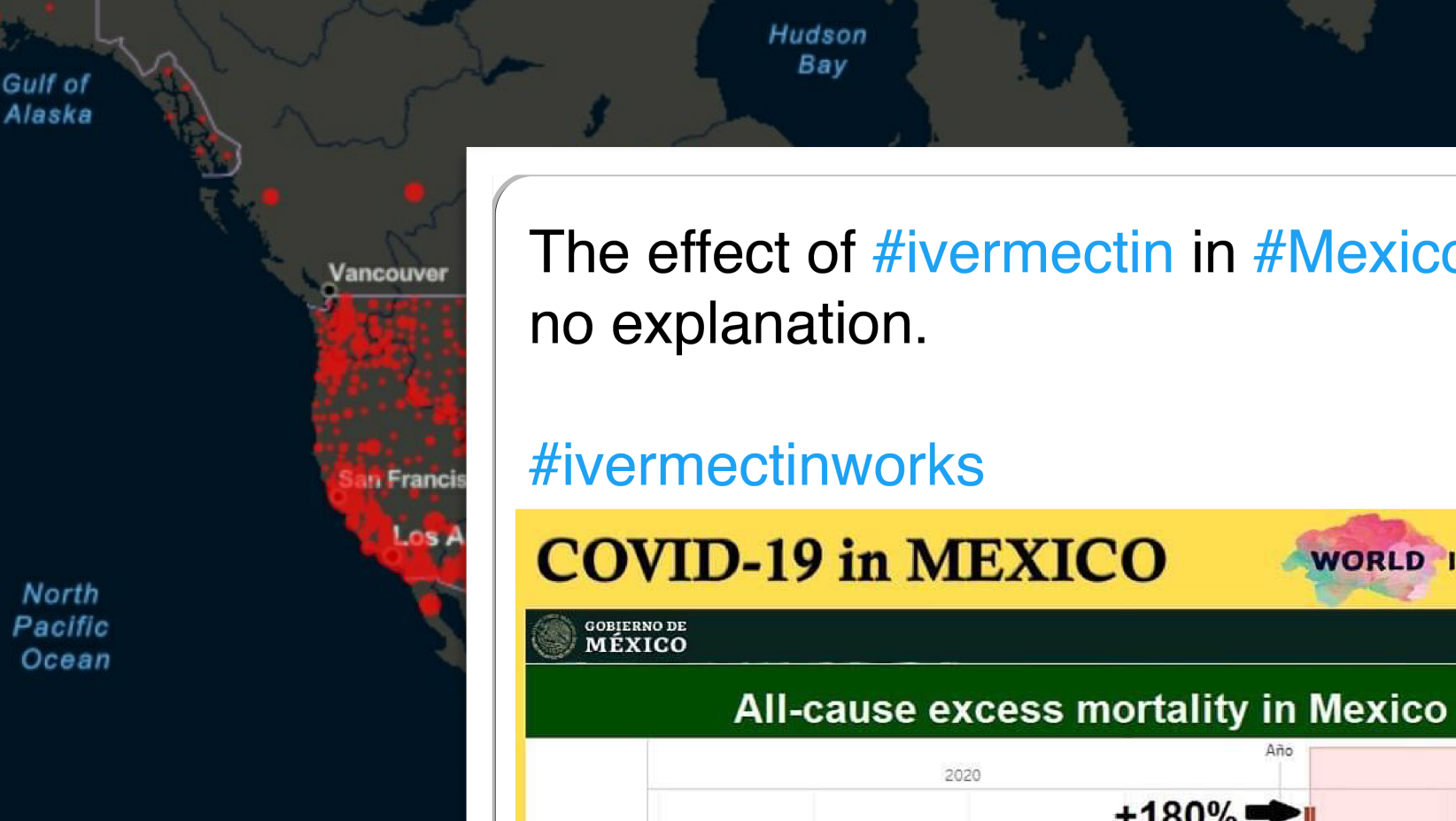
More proof that masks work.....

Kansas COVID-19 7-Day Rolling Average of Daily Cases/Per 100K Population

Mask Counties Vs. No-Mask Mandate Counties
Source: Kansas Department of Health and Environment



We need national leadership...like Canada.
Covid is coded in red.



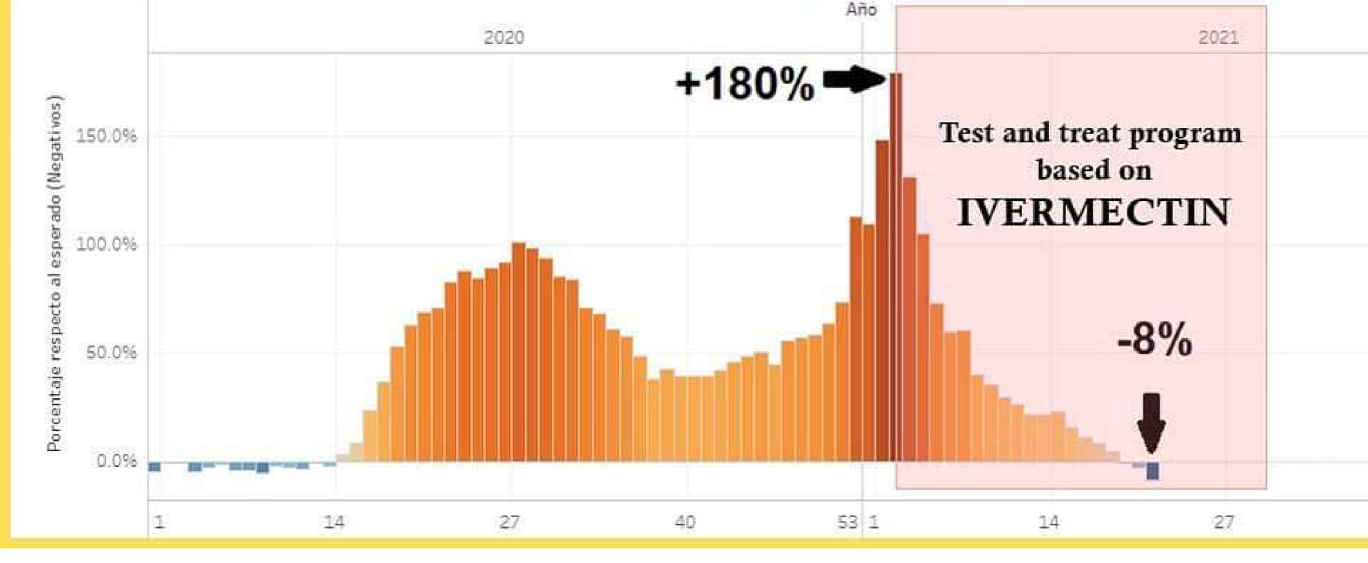
The effect of #ivermectin in #Mexico needs no explanation.

#ivermectinworks

COVID-19 in MEXICO

WORLD IVERMECTIN DAY

All-cause excess mortality in Mexico 2020-2021

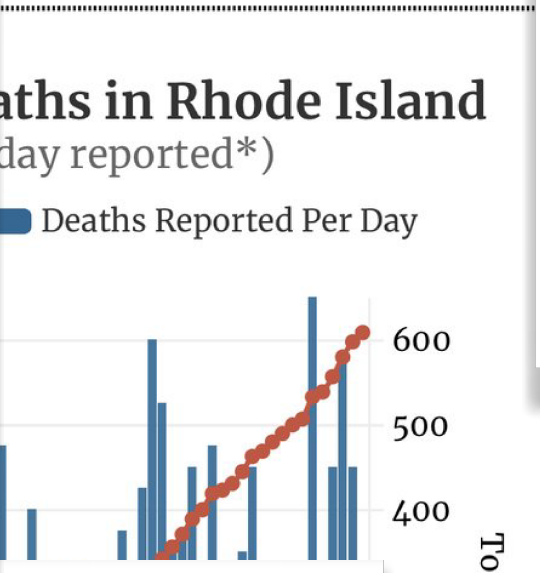


milestone: today Rhode Island's avirus death toll passed 600 people

[wprh.com/news/health/ivermectin-works-in-rhode-island](https://www.wprh.com/news/health/ivermectin-works-in-rhode-island)

Deaths in Rhode Island (day reported*)

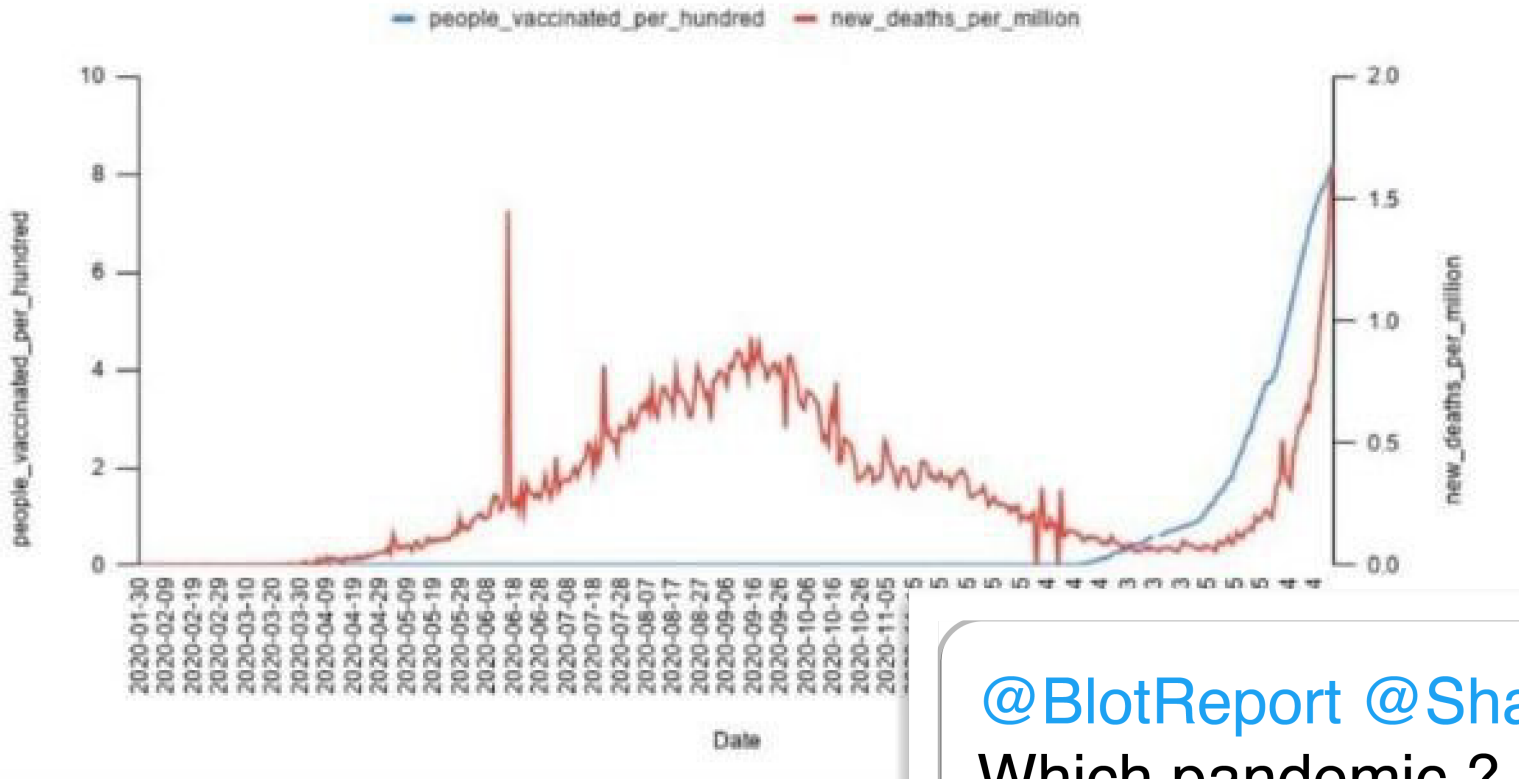
Deaths Reported Per Day



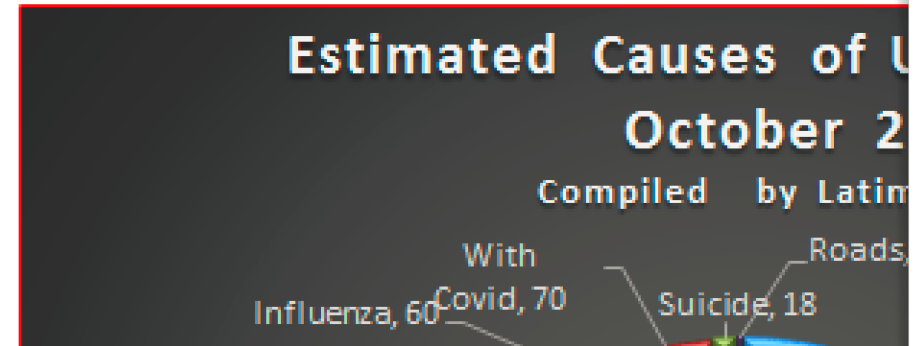
Whoa look what I found, death rates rising in line with vaccination numbers in India.

India - people_vaccinated_per_hundred vs new_deaths_per_million

Source: <https://covid.ourworldindata.org/data/india-covid-data.xlsx>



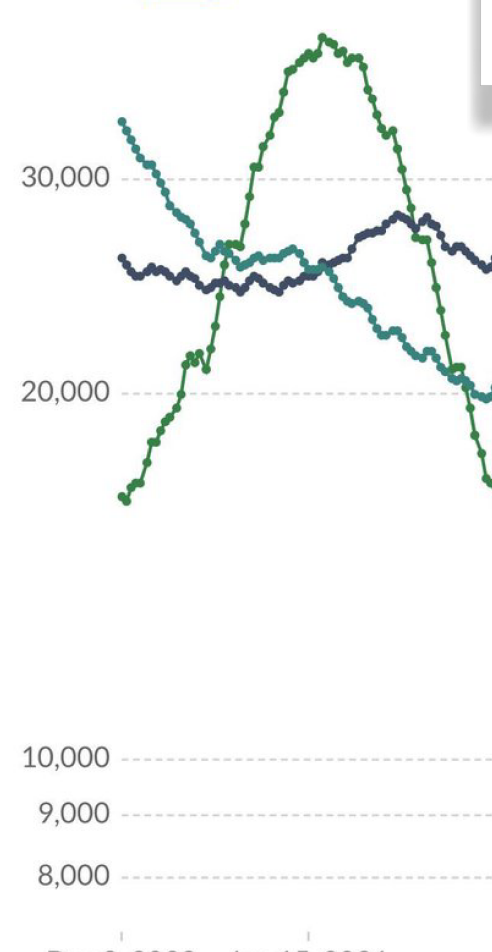
Hot off the press:
just 1 death in 23 today in UK w



Vaccines work.

Number of COVID-19 patients in hospital

LINEAR LOG + Add country

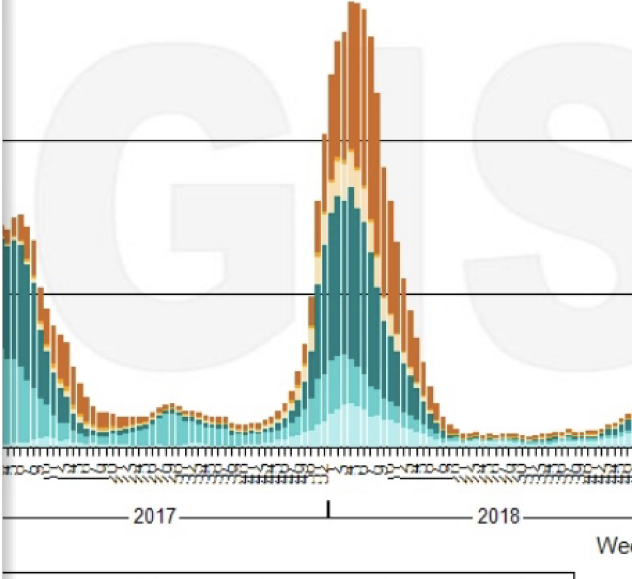


Laboratory Surveillance Information

Global Influenza Surveillance and Response System (GISRS)

Evolution of influenza viruses

Number of specimens positive for influenza by subtype



Two words

Ivermectin works

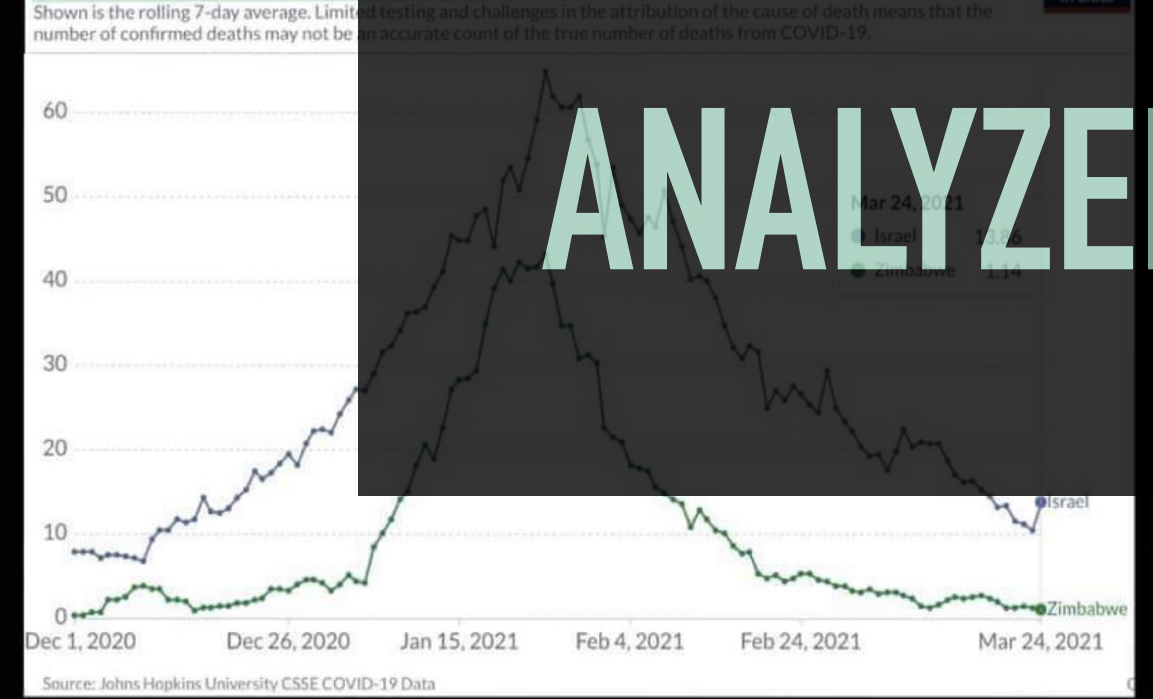
#ivermectinworks

ISRAEL v. ZIMBABWE

Israel is fighting with mass vaccination, Zimbabwe with mass Ivermectin (approved 26th Jan)

Daily new confirmed COVID-19 deaths

Source: Johns Hopkins University CSSE COVID-19 Data

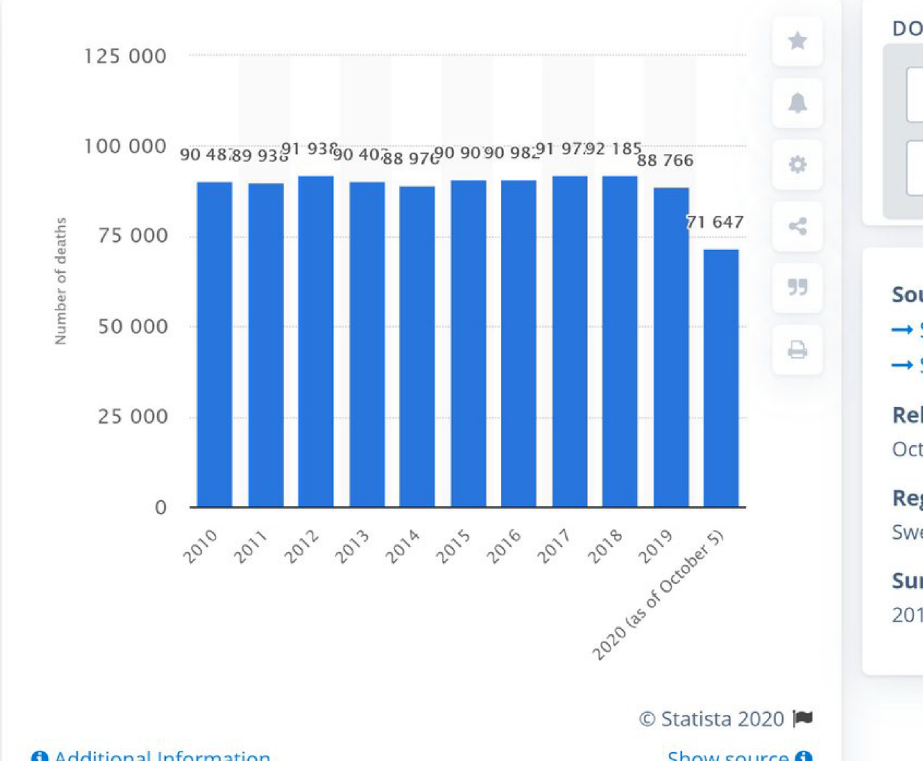


@BlotReport @ShawnG927 @ici_cana

Which pandemic ? Nothing to see here

Society + Demographics

Number of deaths in Sweden from 2010 to 2020

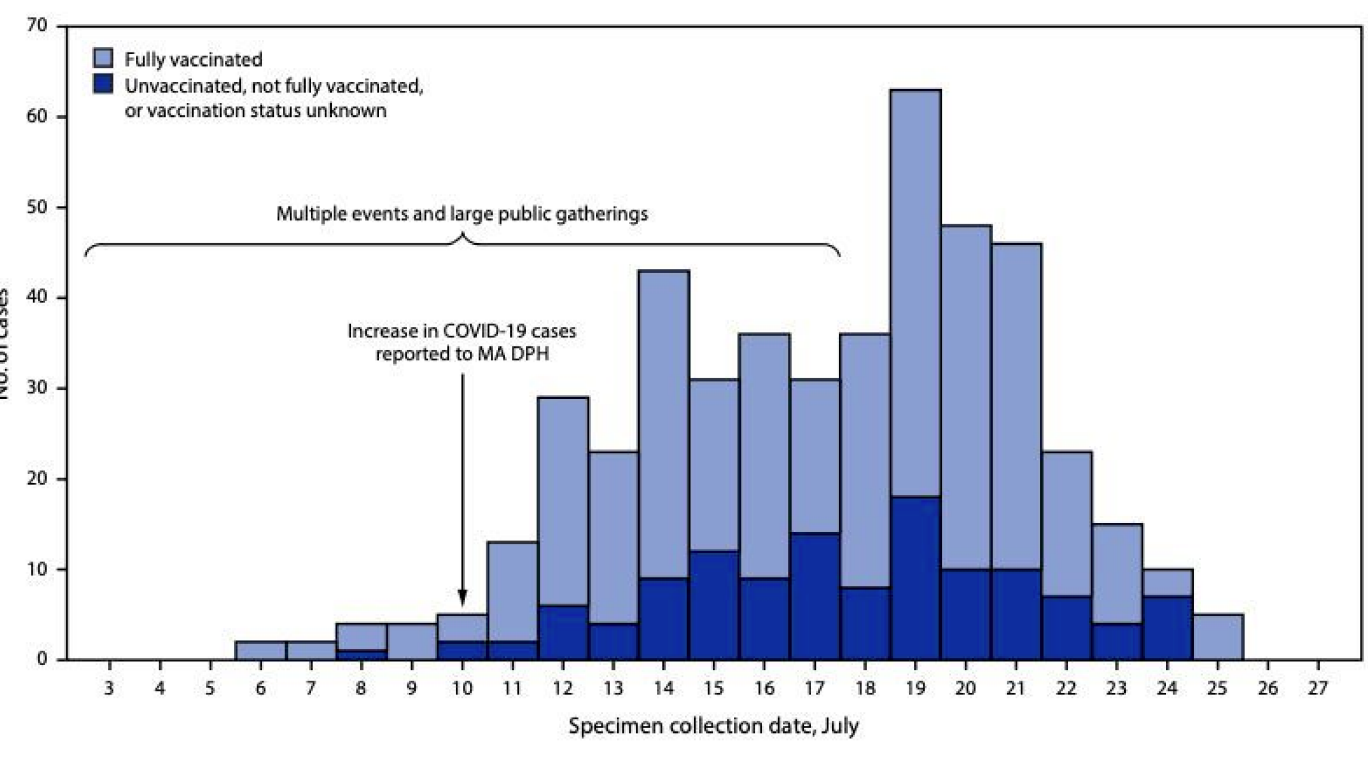


According to official data, the vaccinated are the super-spreaders.

<https://cdc.gov/mmwr/volumes/7...>

<https://t.me/EARTH20GENESIS...>

FIGURE 1. SARS-CoV-2 infections (N = 469) associated with large public gatherings, by date of specimen collection and vaccination status* — Barnstable County, Massachusetts, July 2021



ccinated HT @FungalDoc & Reddit.

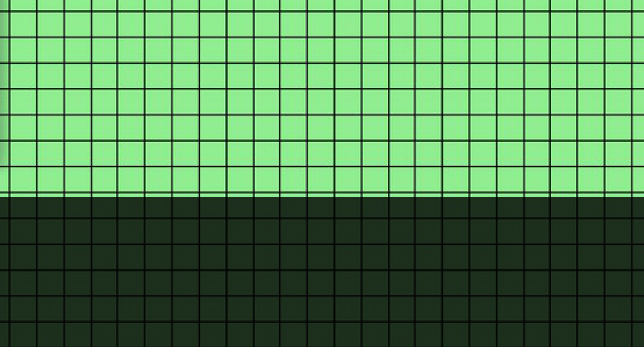
Automatic breakthrough COVID-19 infections

Currently...

For every 102K vaccinated

There are 100 symptomatic breakthrough infections

And ONE death

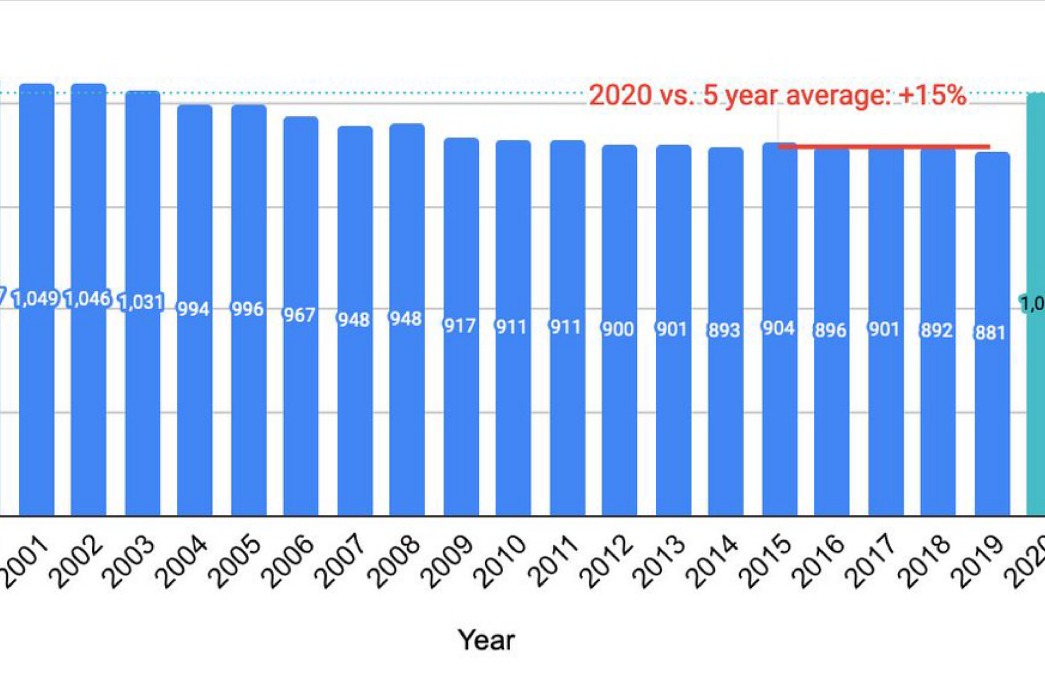


acc mortality in the US is as bad as it was

#Covid #Corona #Coronavirus

usted Mortality Rate [United States]

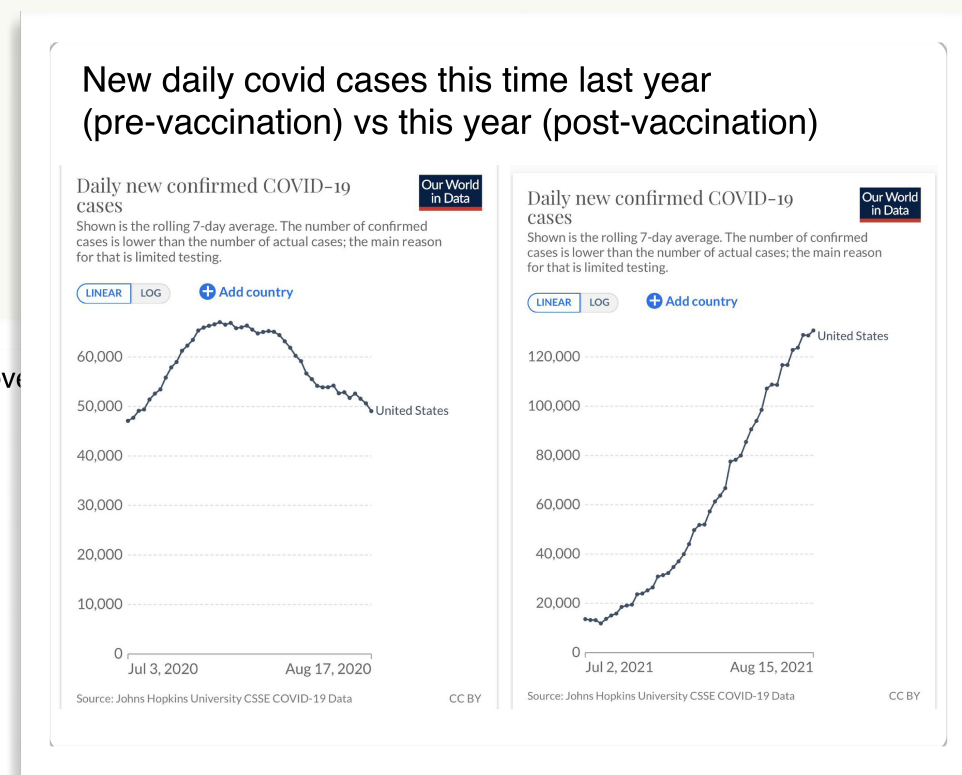
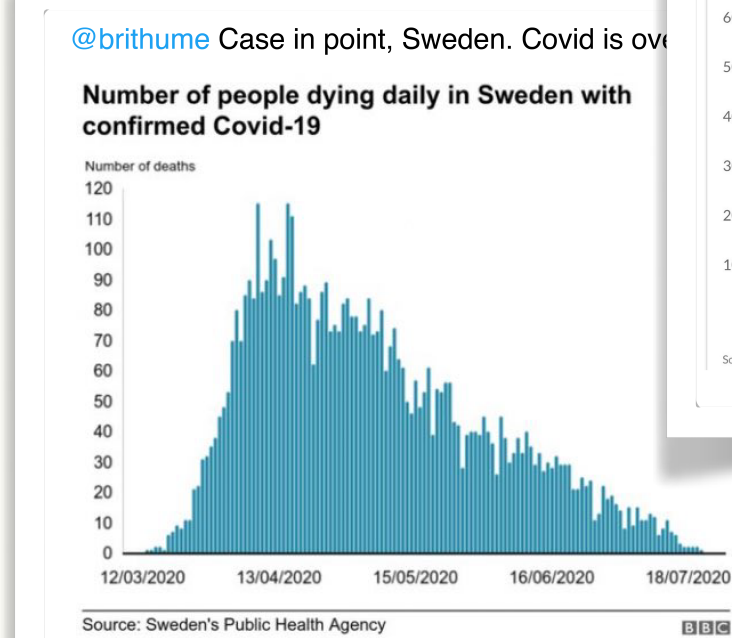
00; Adjusted for Age Distribution (United States 2020)



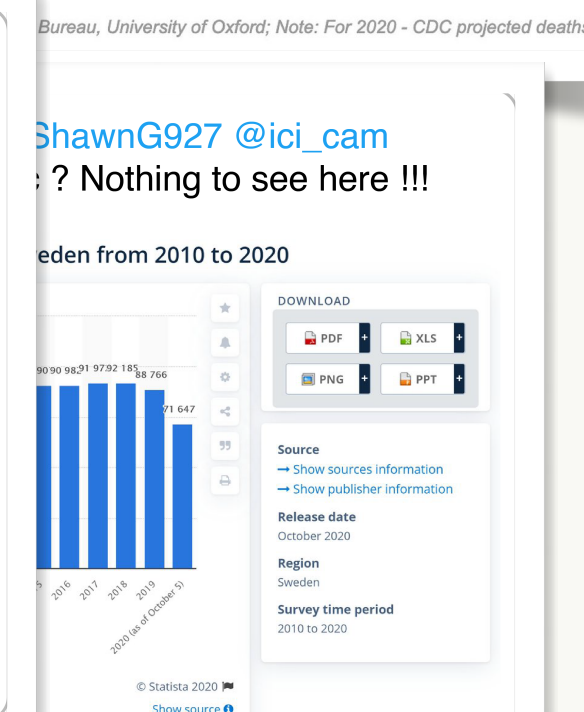
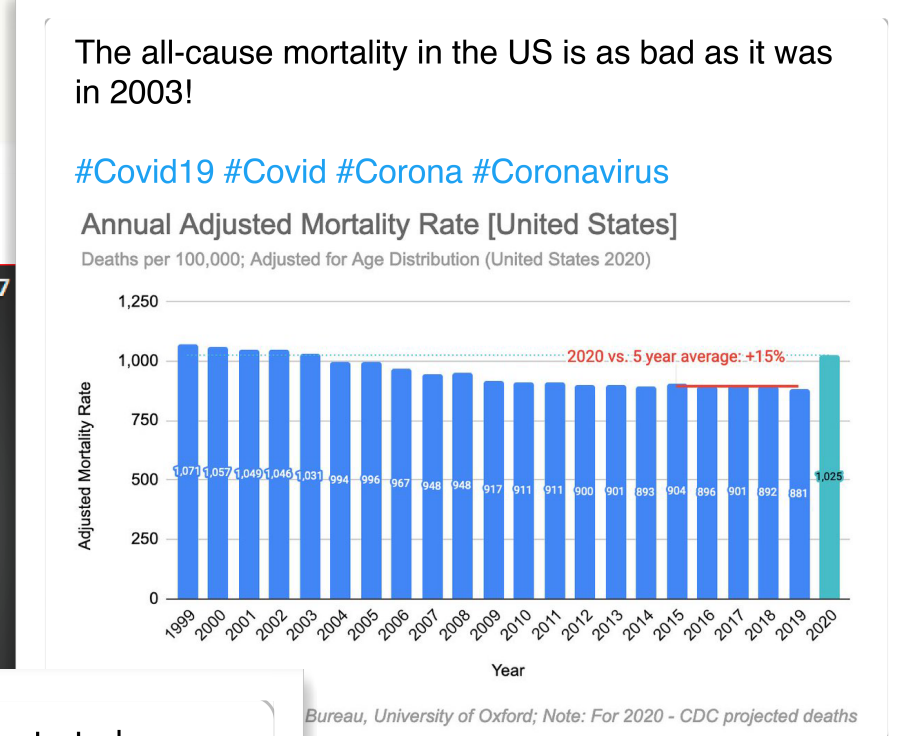
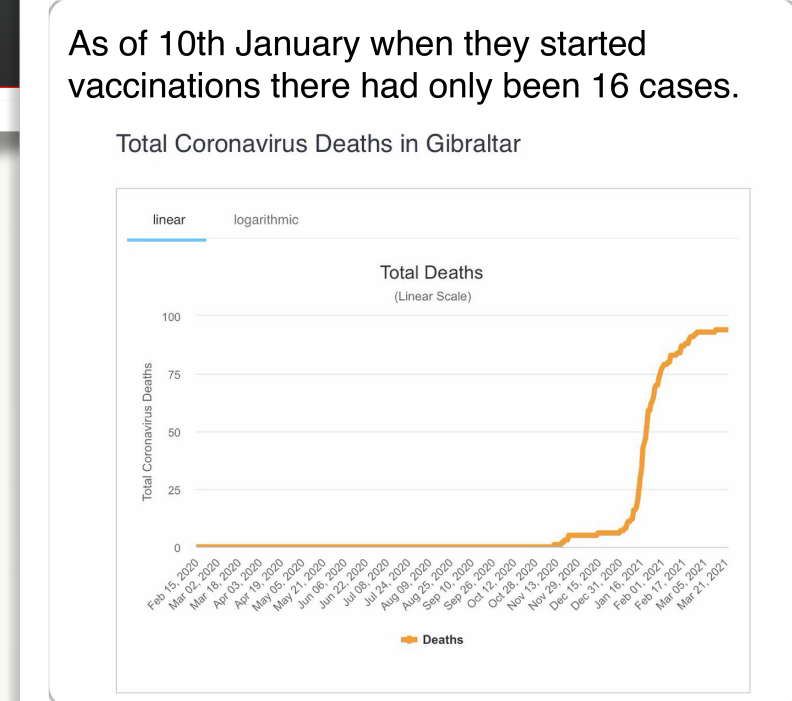
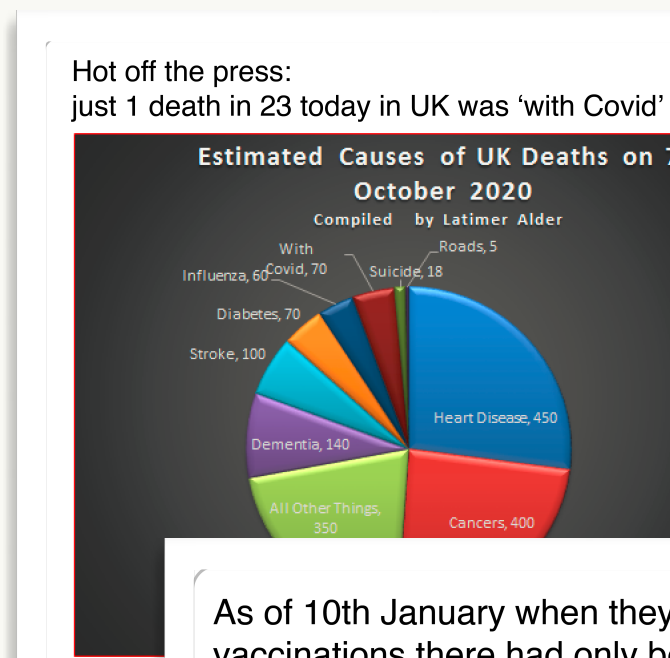
S. Census Bureau, University of Oxford; Note: For 2020 - CDC projected deaths

ANALYZED 10K COVID-19 VISUALIZATIONS SHARED ON TWITTER

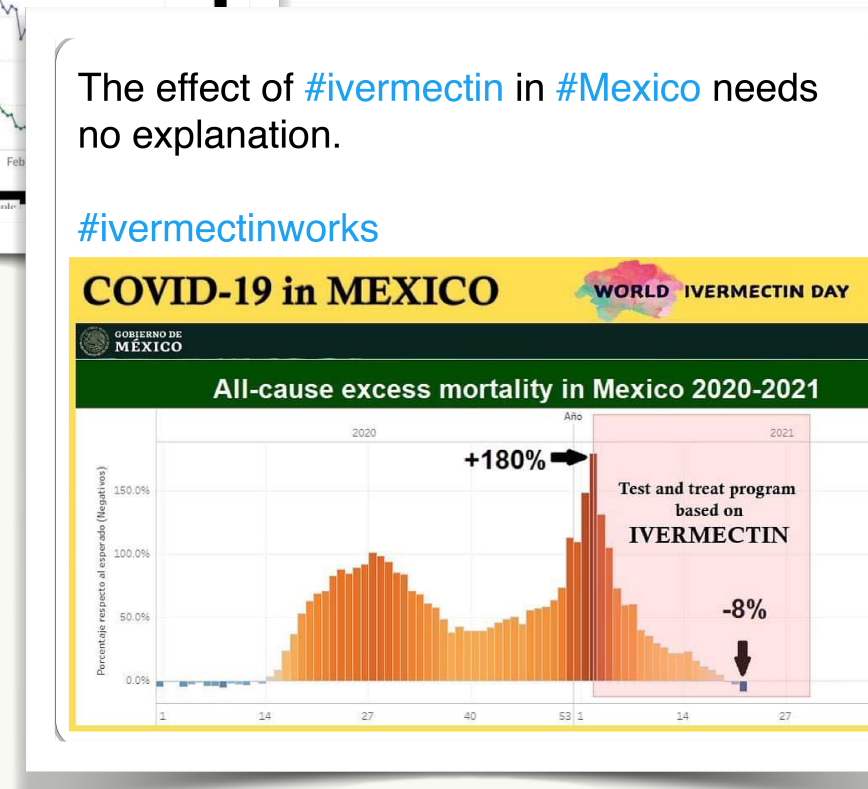
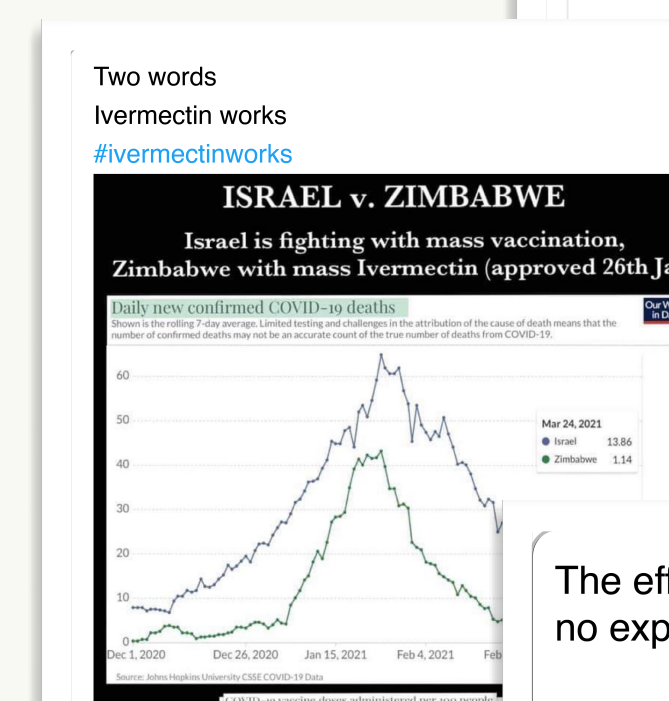
Cherry-picking



Arbitrary threshold



Causal inference

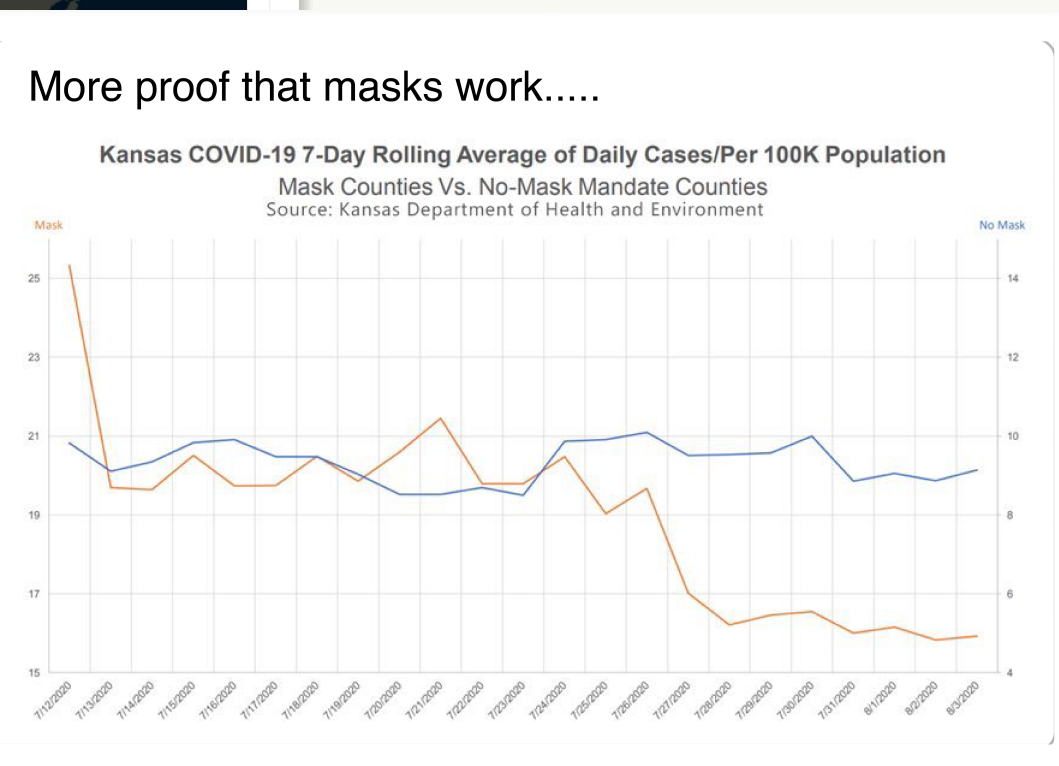


Issues with data validity

We need national leadership...like Canada. Covid is coded in red.



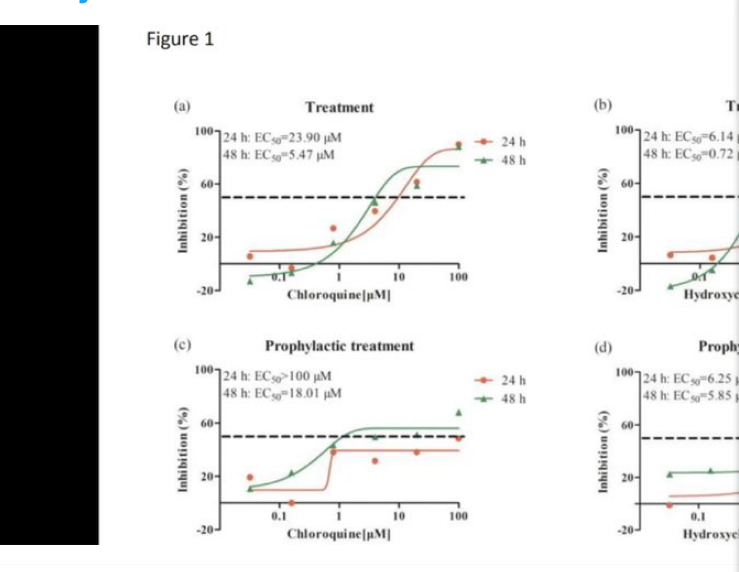
Incorrect reading



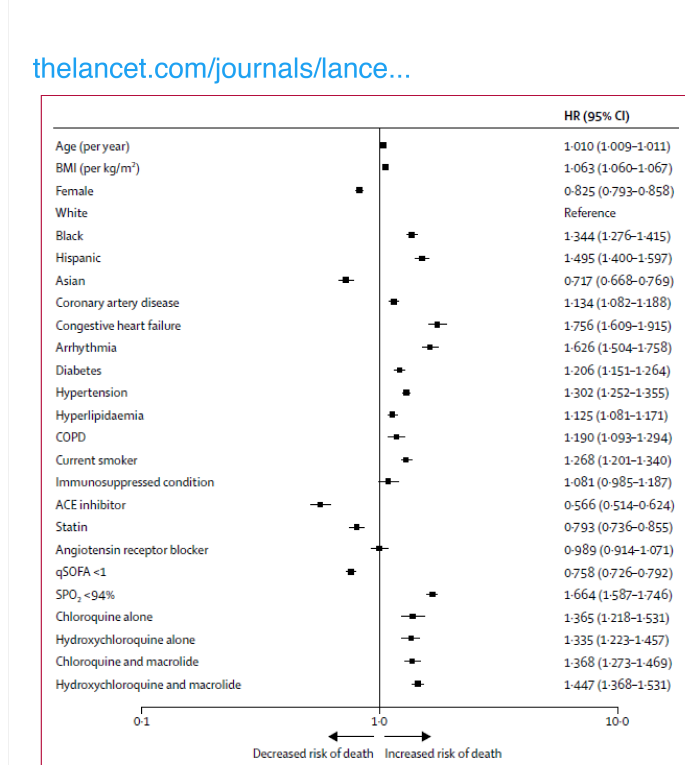
Misrepresentation of scientific results

Hydroxychloroquine probably better

m.youtube.com/watch?feature=...



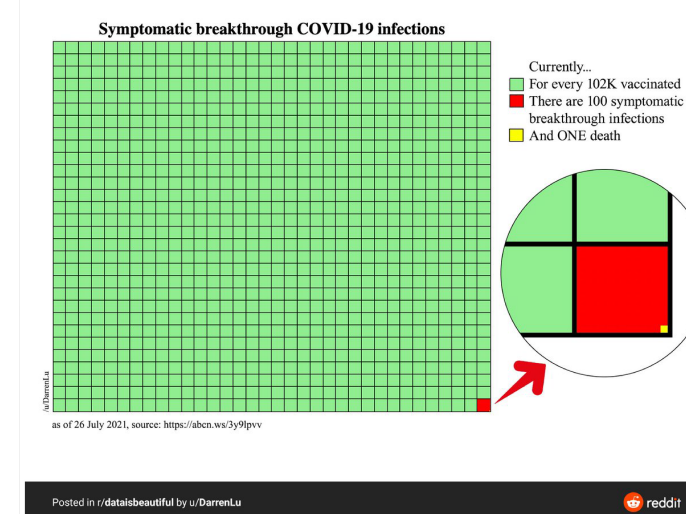
Taking HCO is as strongly associated with increased coronavirus death risk as DIABETES.



Statistical nuance

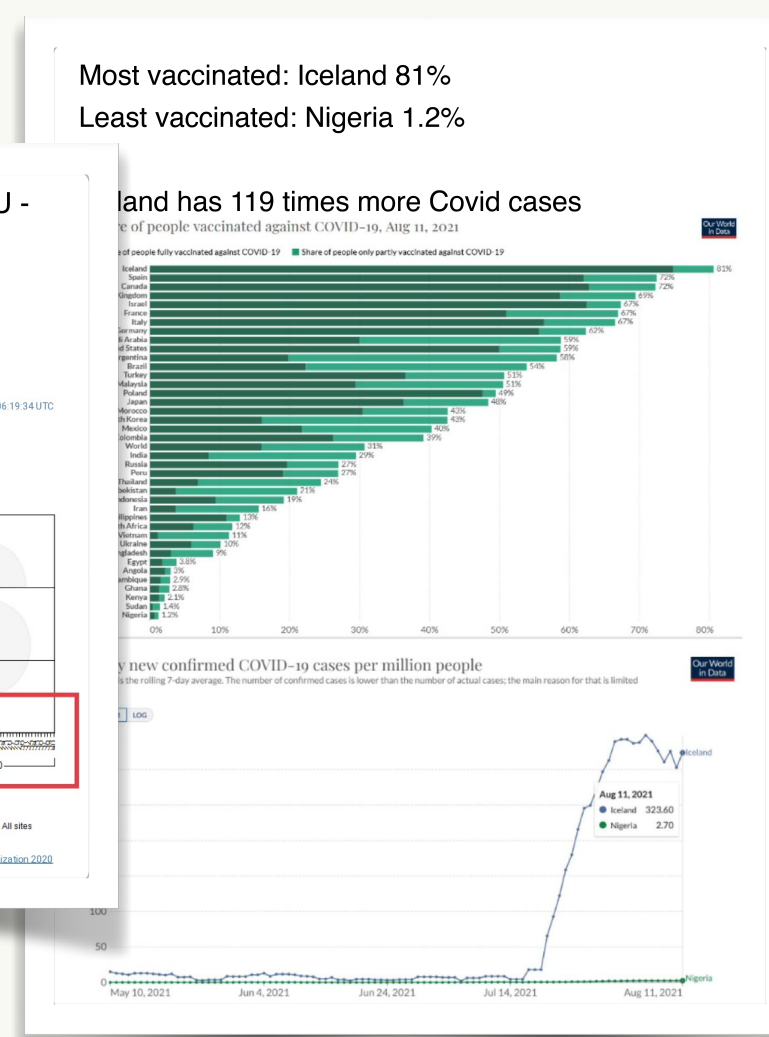
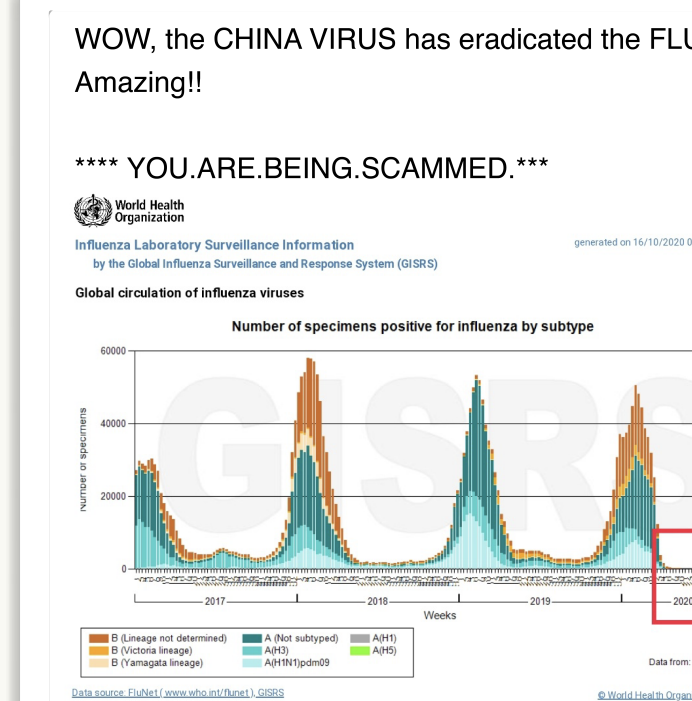
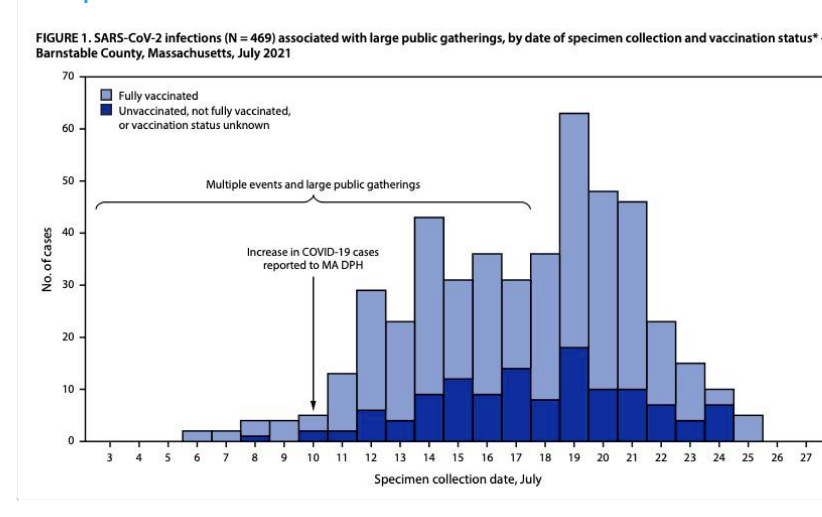
Make this go viral. Vaccines work.

#GetVaccinated HT @FungalDoc & Reddit.



According to official data, the vaccinated are the super-spreaders.

<https://cdc.gov/mmwr/volumes/7...>
<https://t.me/EARTH20GENESIS...>



The effect of [#ivermectin](#) in [#Mexico](#) needs no explanation.

[#ivermectinworks](#)



Causal
inference

Screenshot of a government dashboard

Sharp mortality drop is a salient visual feature

Added annotation assigns a cause-and-effect

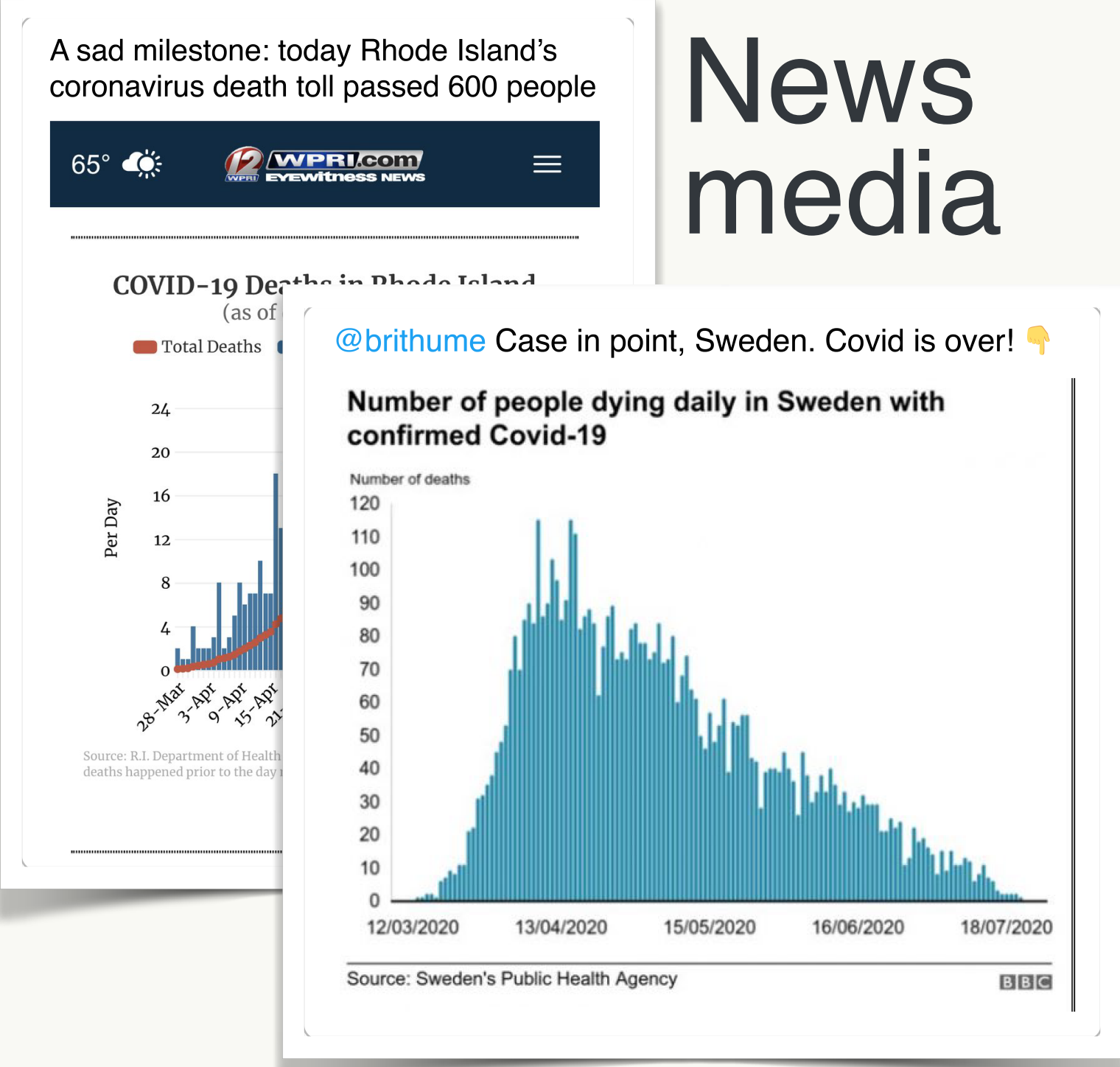
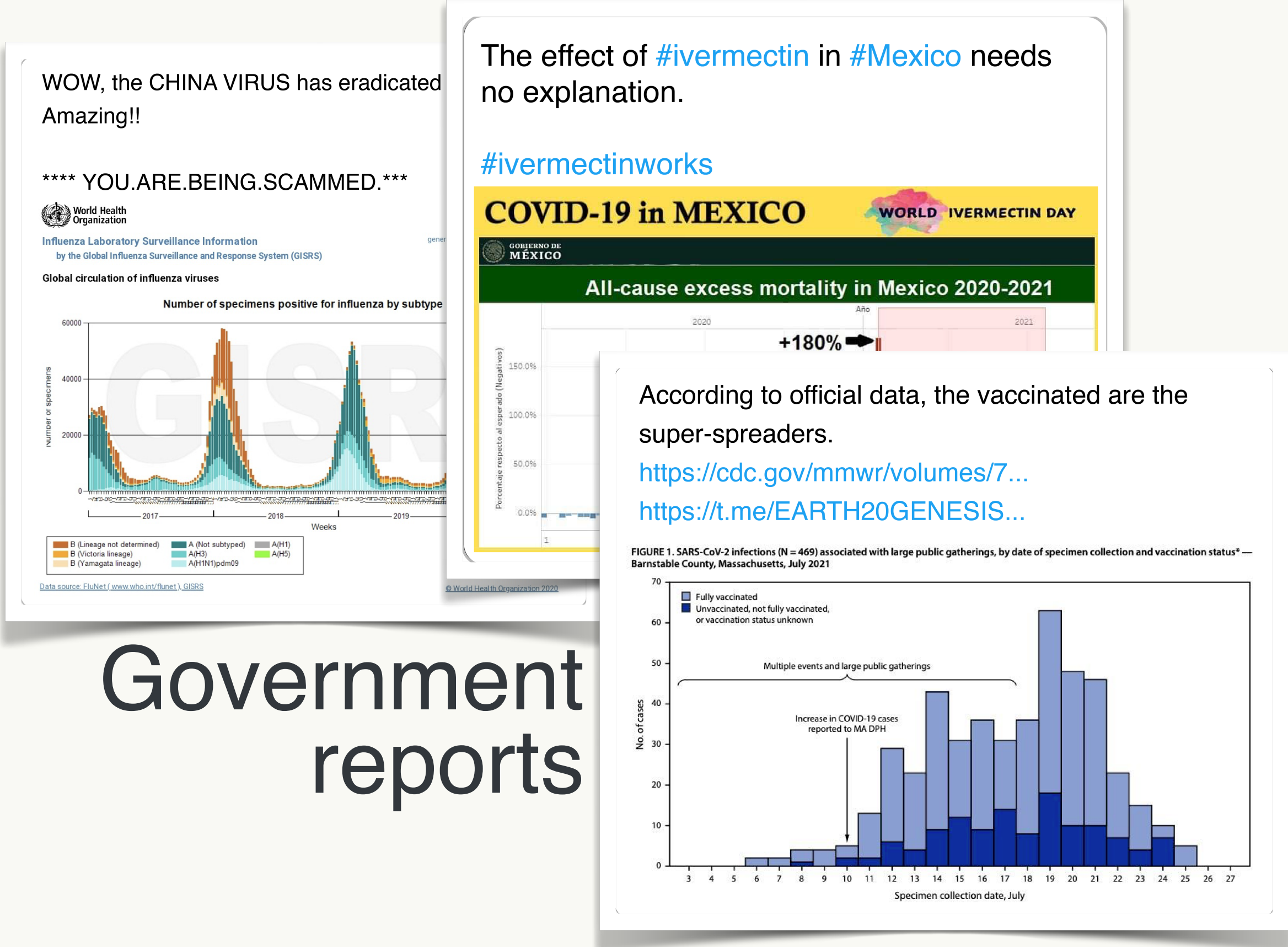
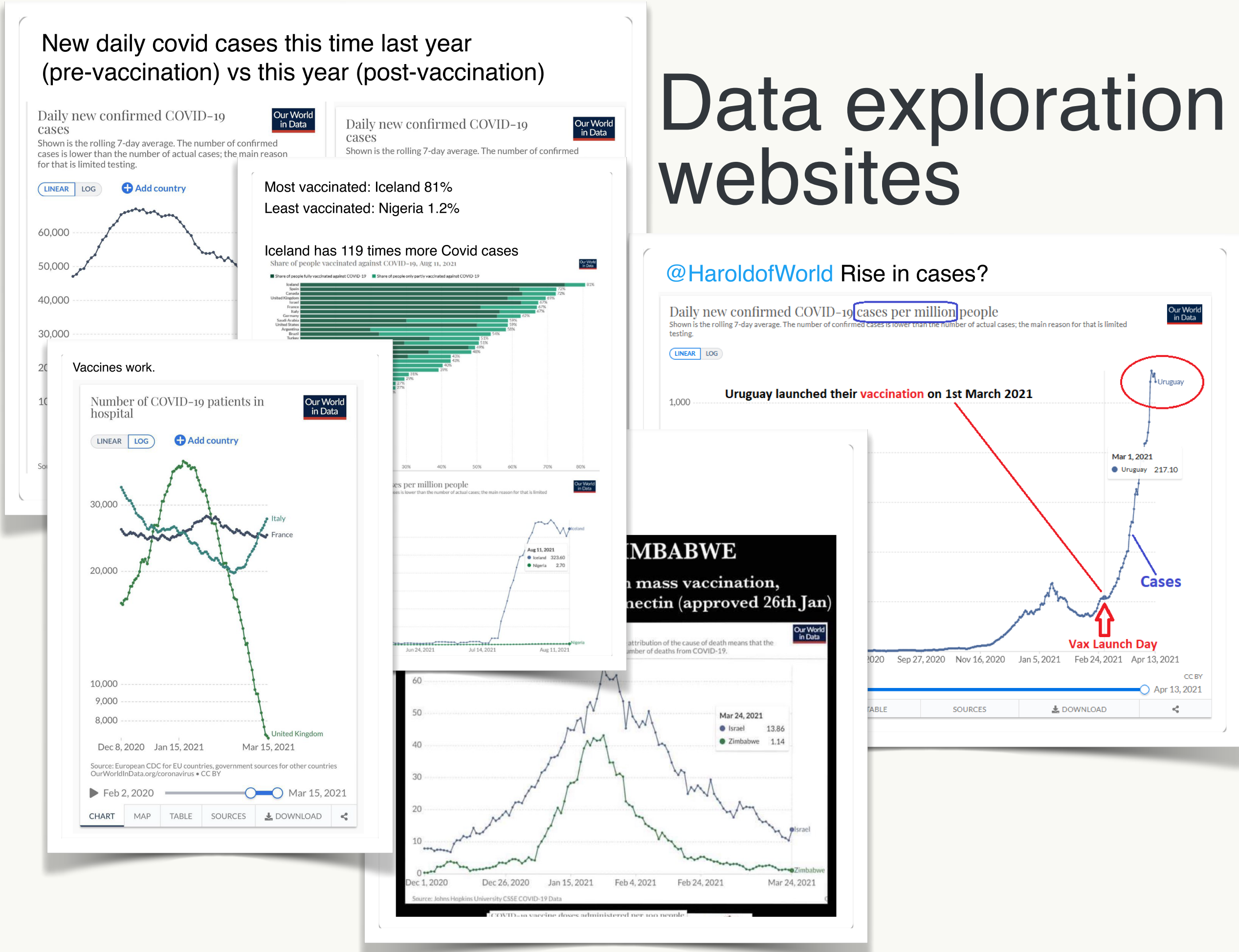
Tweet text further explains the argument

TAKEAWAYS

89% of charts
with reasoning errors
**do not violate any
visualization guidelines**

The majority of **misleading
charts** are screenshots
from **reputable sources**

Sources of Deceptive Visualizations



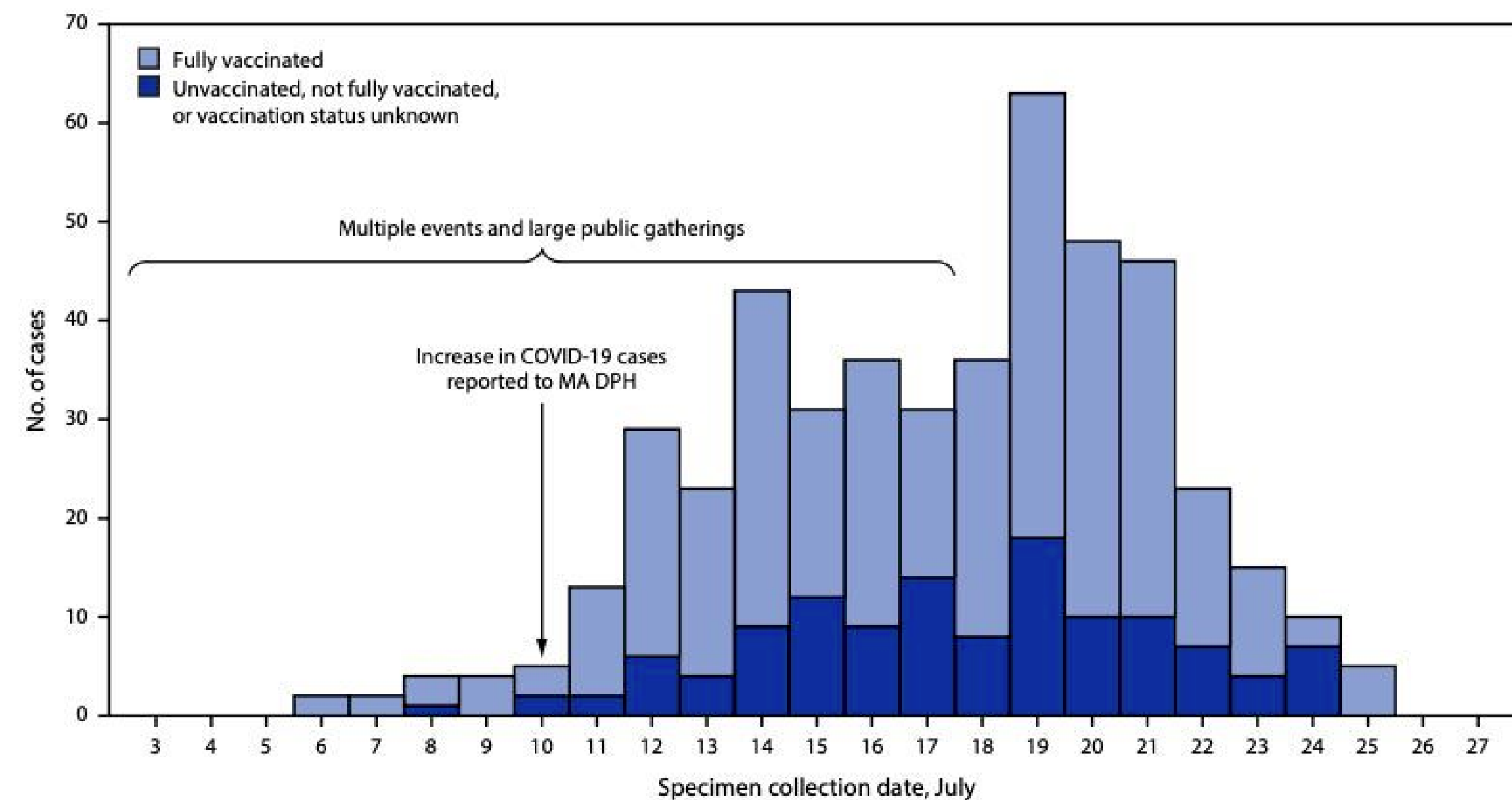
VULNERABLE VISUALIZATIONS

According to official data, the vaccinated are the super-spreaders.

<https://cdc.gov/mmwr/volumes/7...>

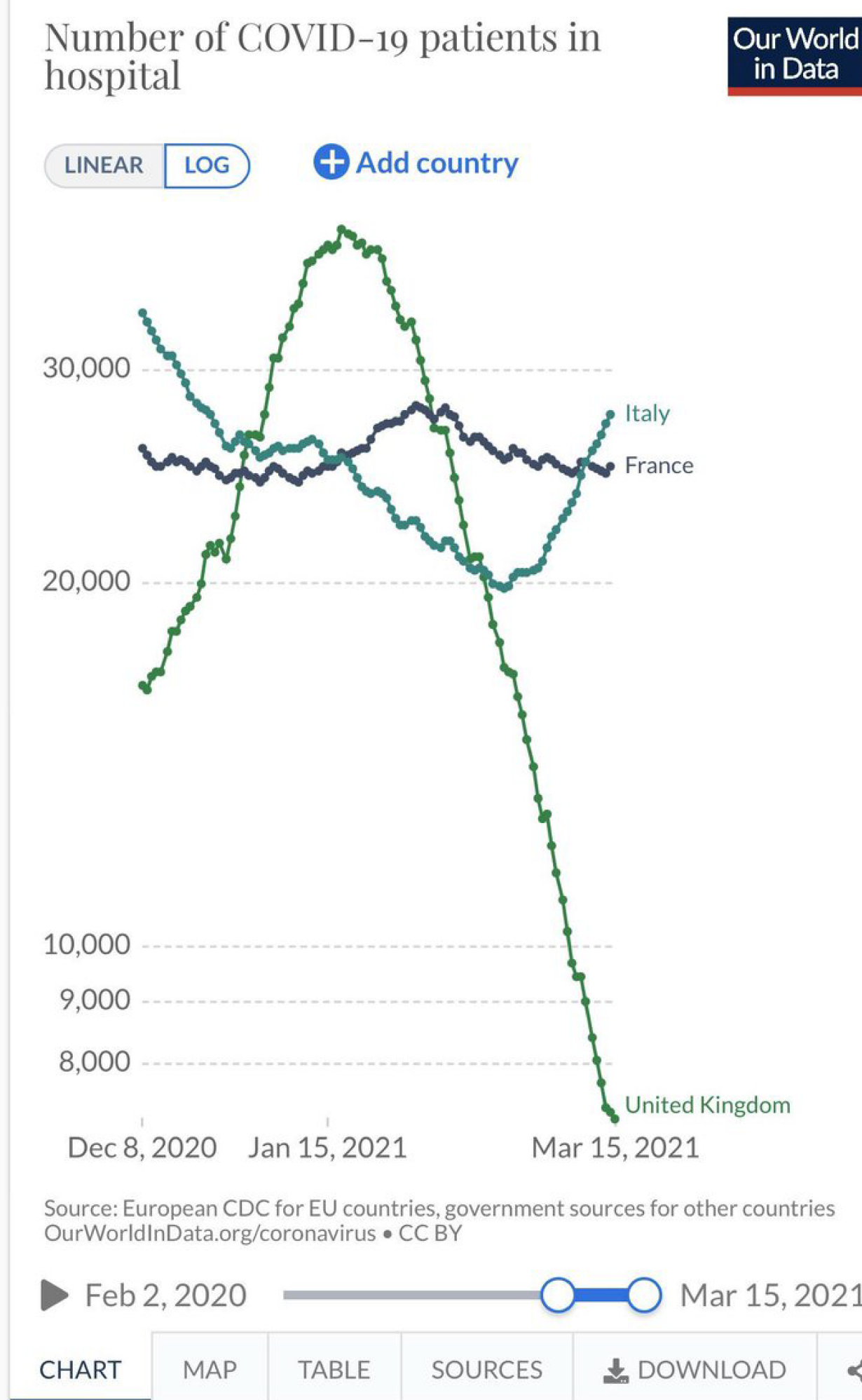
<https://t.me/EARTH20GENESIS...>

FIGURE 1. SARS-CoV-2 infections (N = 469) associated with large public gatherings, by date of specimen collection and vaccination status* — Barnstable County, Massachusetts, July 2021



Notes outside the chart

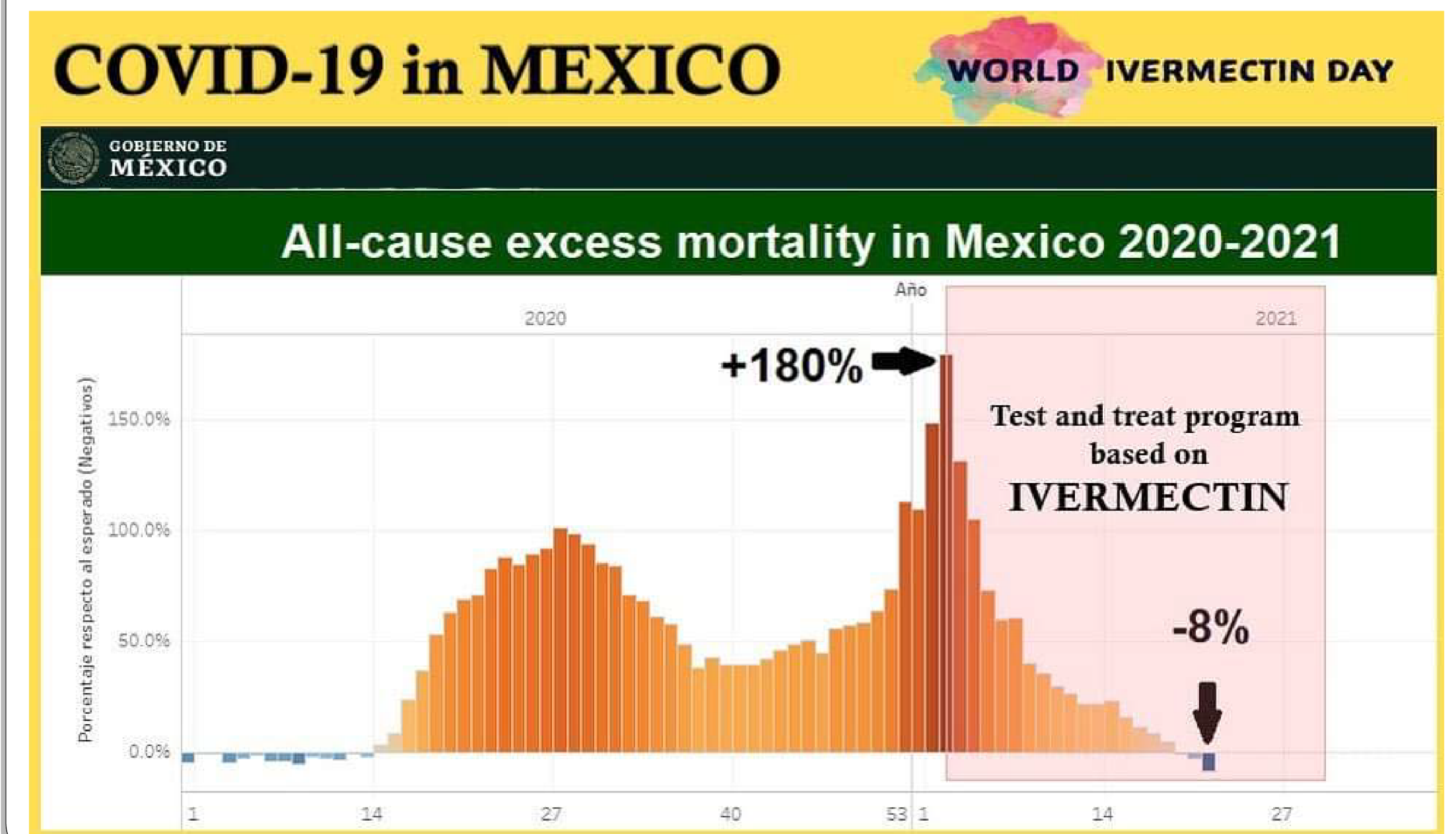
Vaccines work.



Unrestricted interaction

The effect of [#ivermectin](#) in [#Mexico](#) needs no explanation.

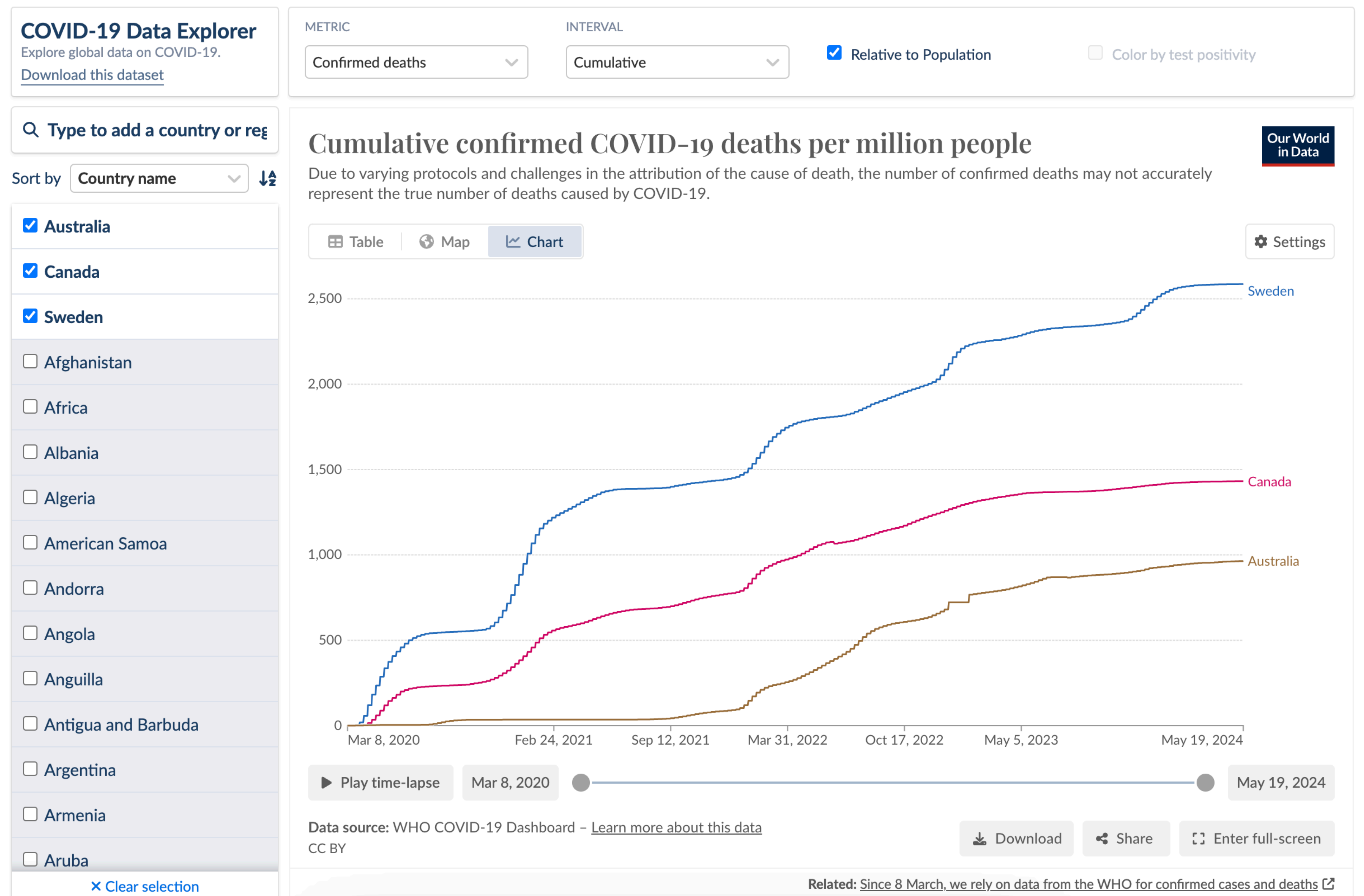
[#ivermectinworks](#)



Added annotations

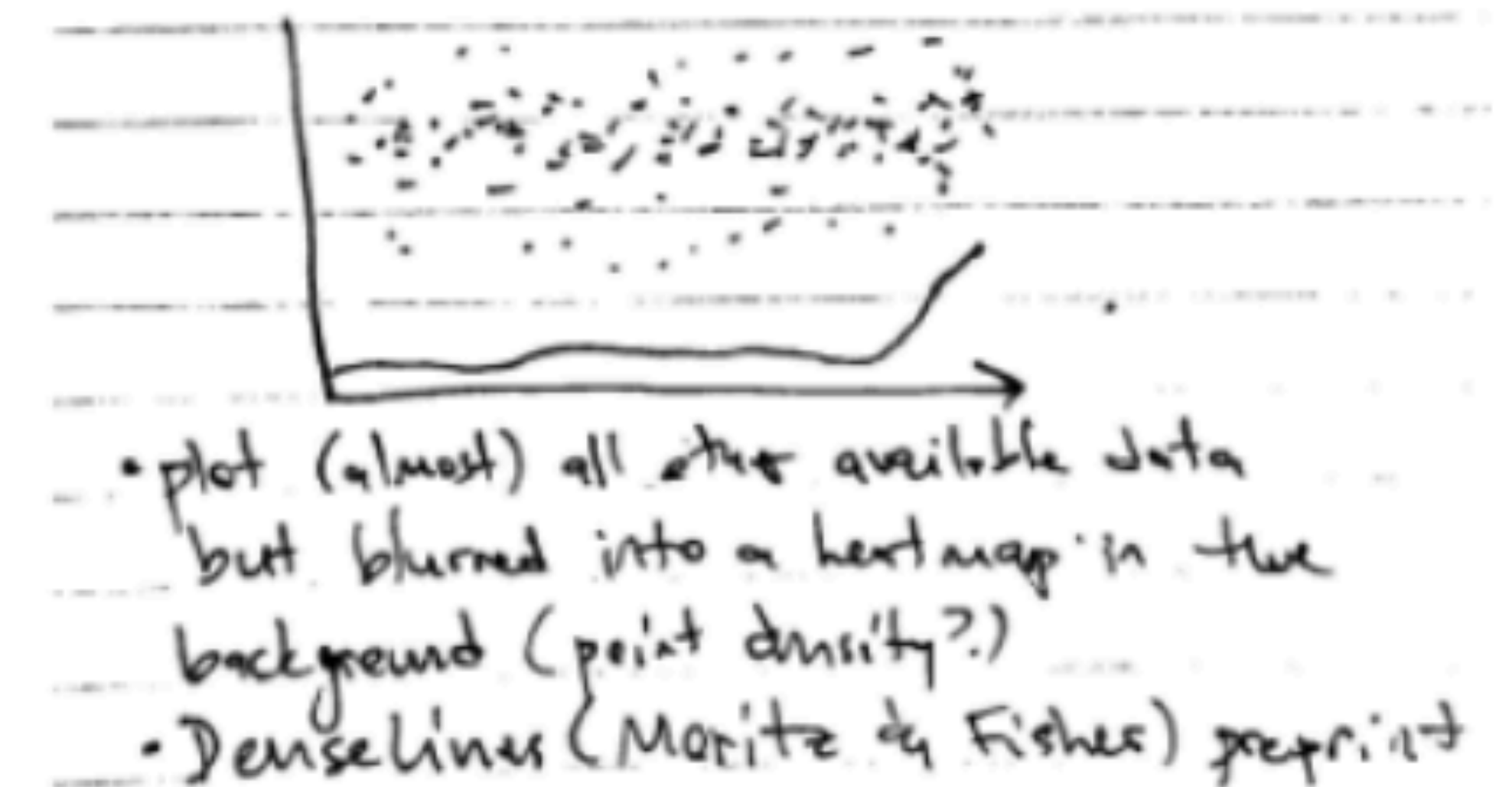
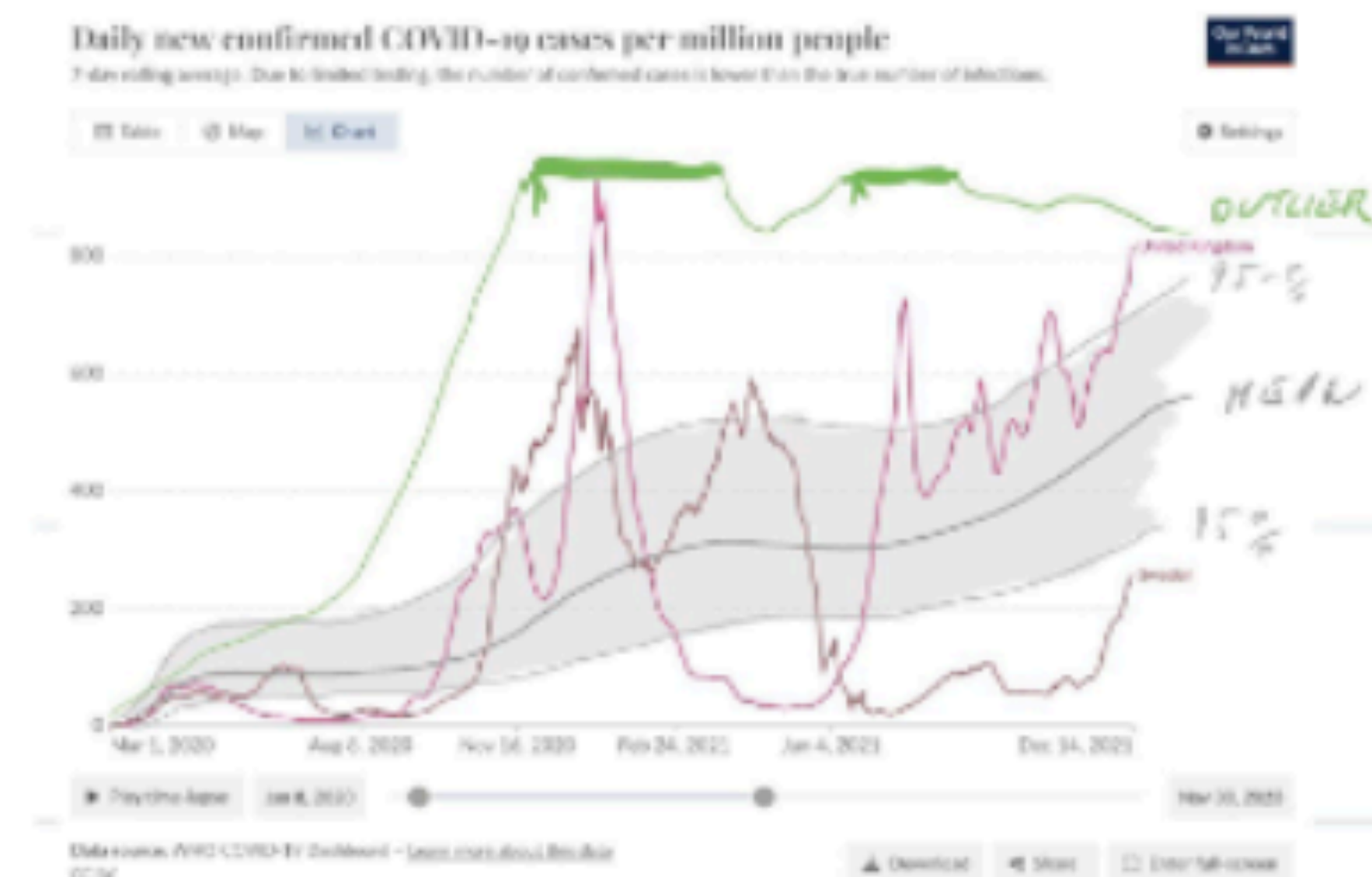
PROTECTING VISUALIZATIONS

Can we design data exploration tools that prevent cherry-picking while maintaining freedom of exploration?

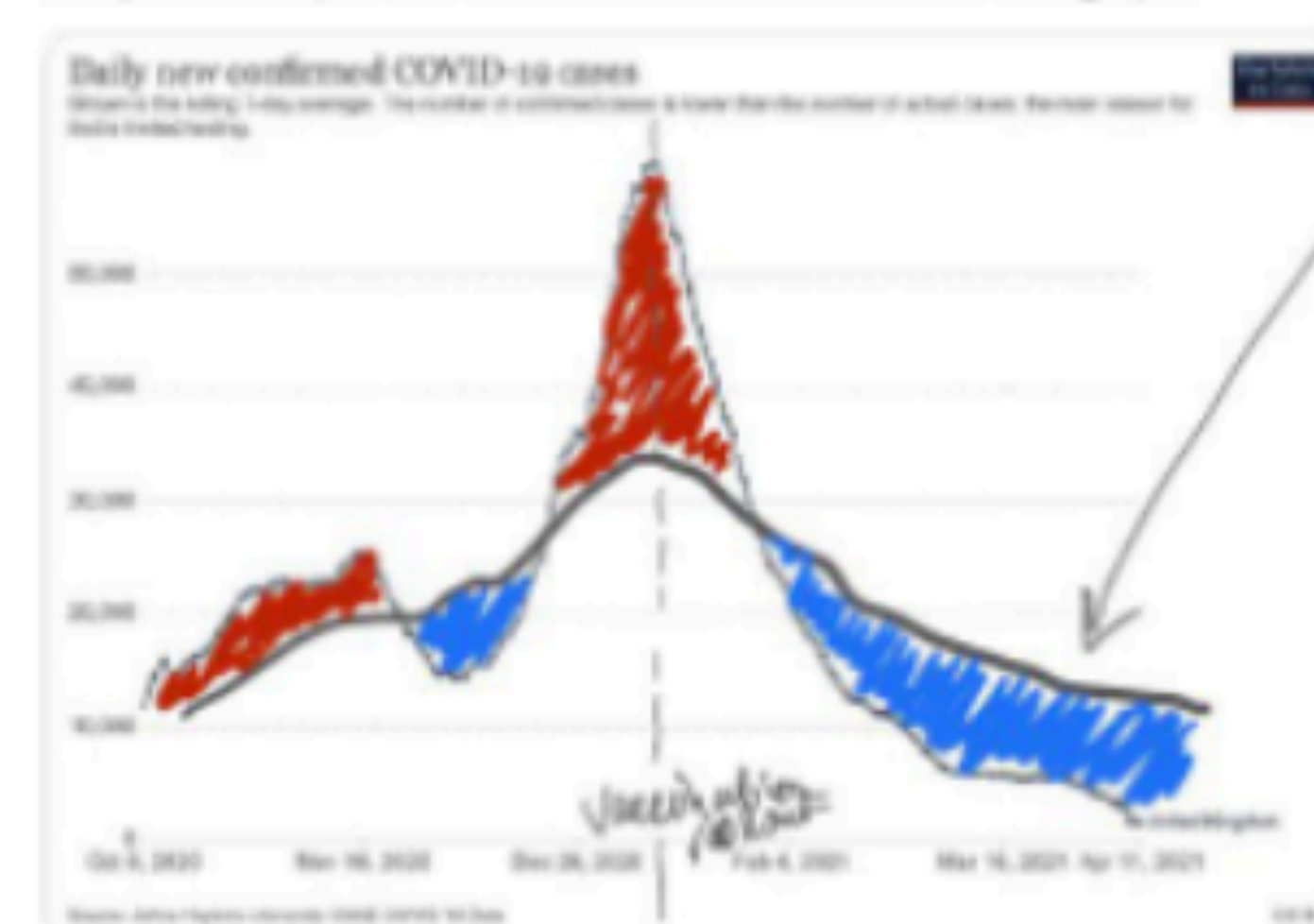


DESIGNING GUARDRAILS

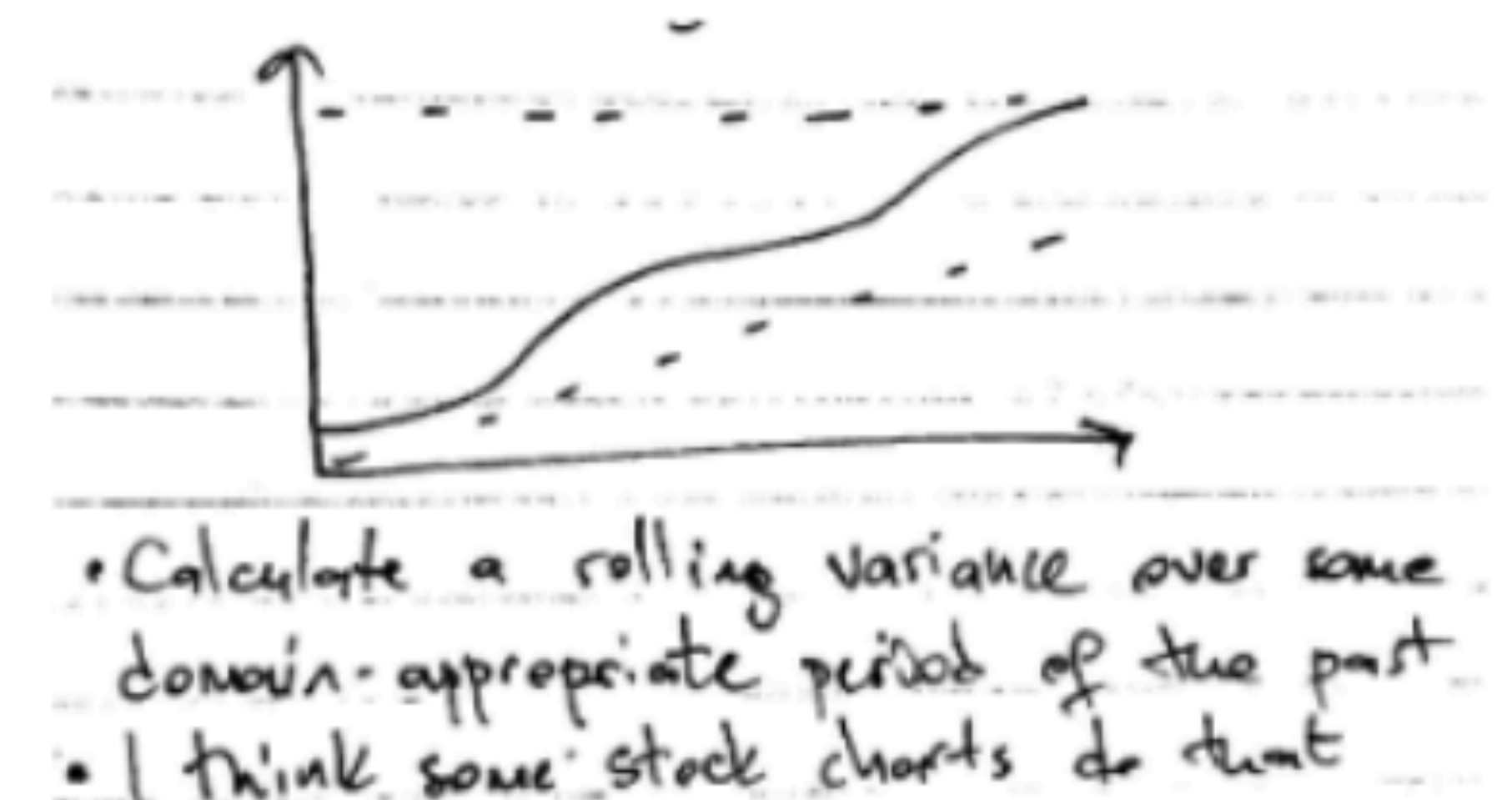
Parallel Design by 4 Authors Cheery-Picking in Temporal Charts (Stocks, COVID)



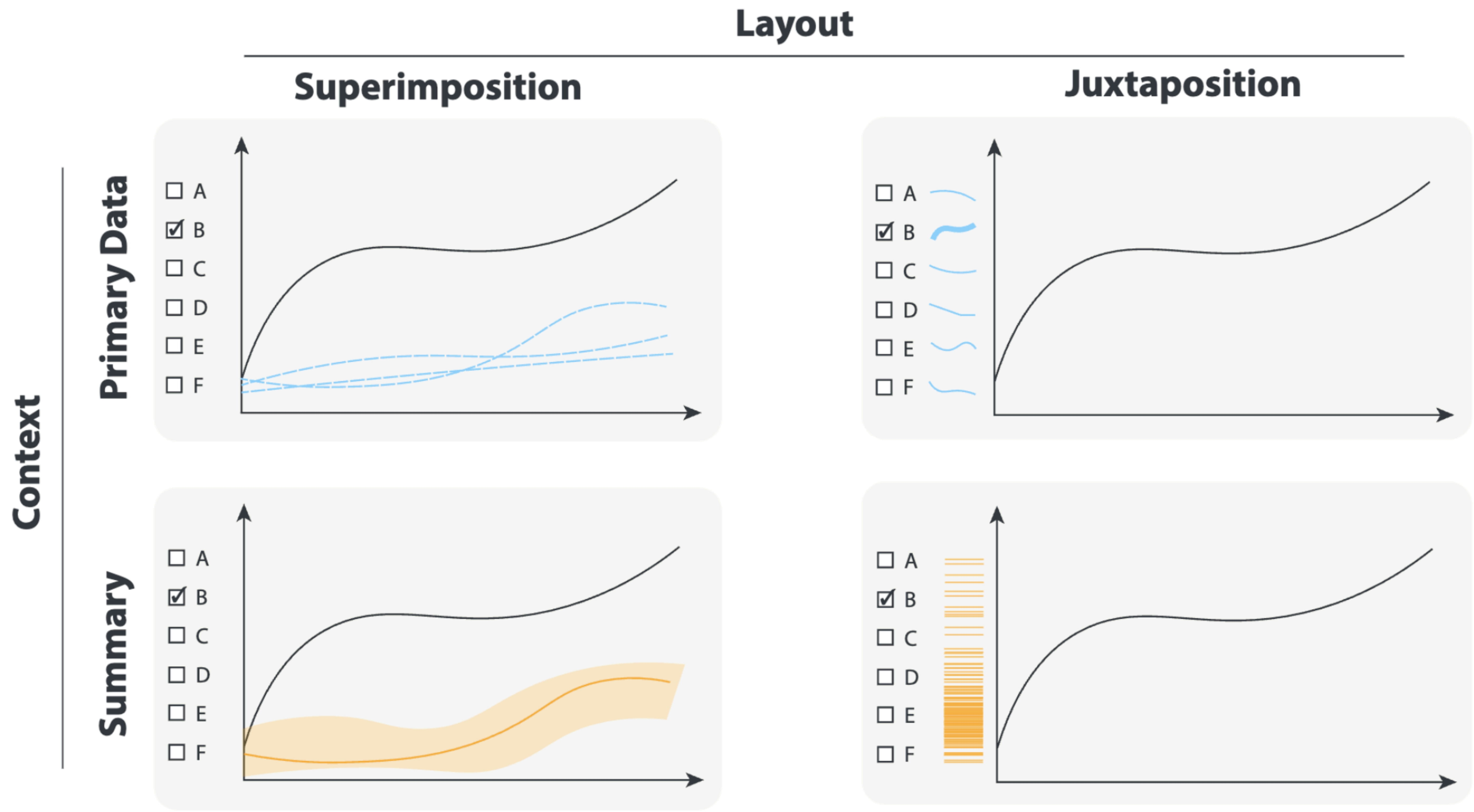
The positive impact of the UK's vaccination efforts in one graph:



some sort of weighted avg for similar population countries, all aligned around their vaccination rollout date (to model UK's in relative error) the avg is also very problematic (outliers, applies a weight) maybe several lines instead, forming some sort of an envelope?



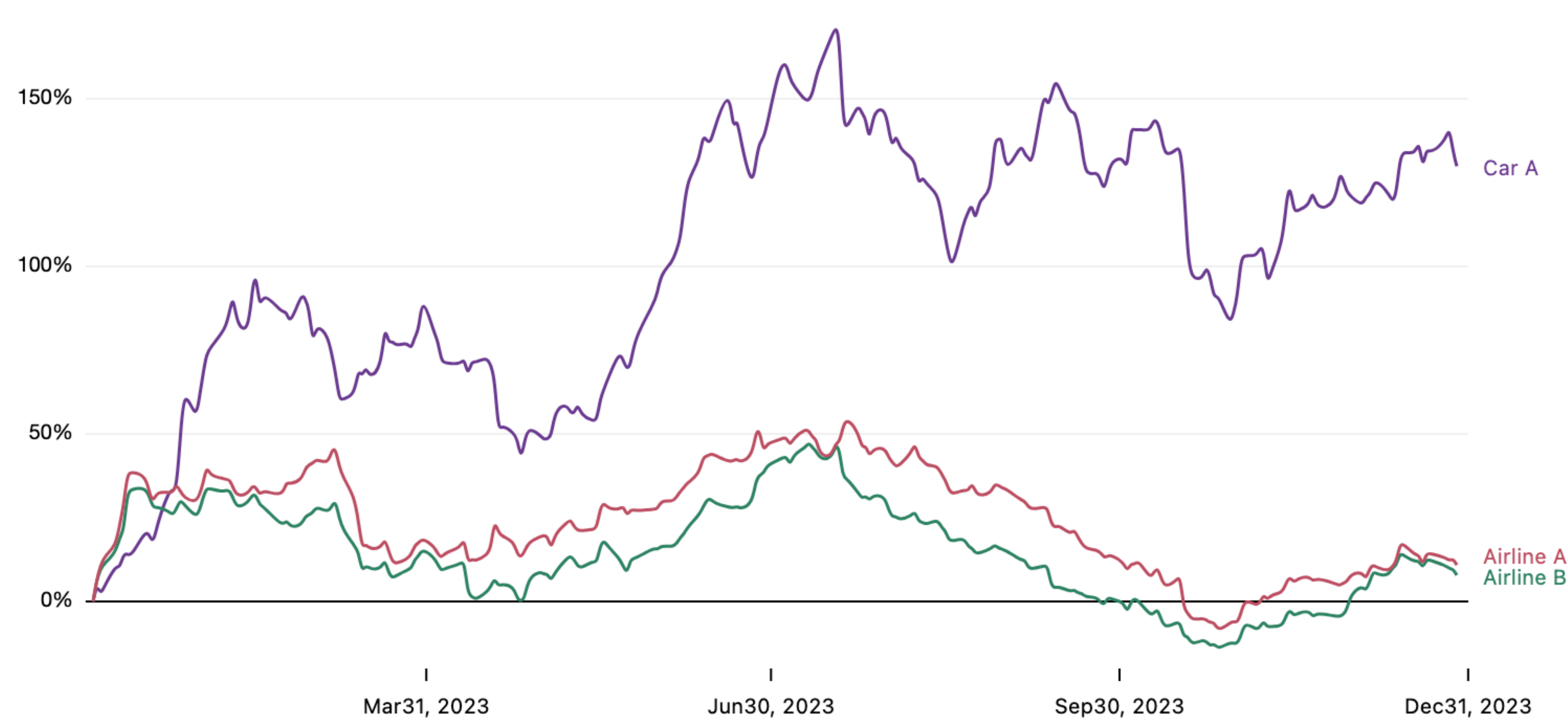
DESIGN SPACE



CHERRY-PICKING: NO GUARDRAILS

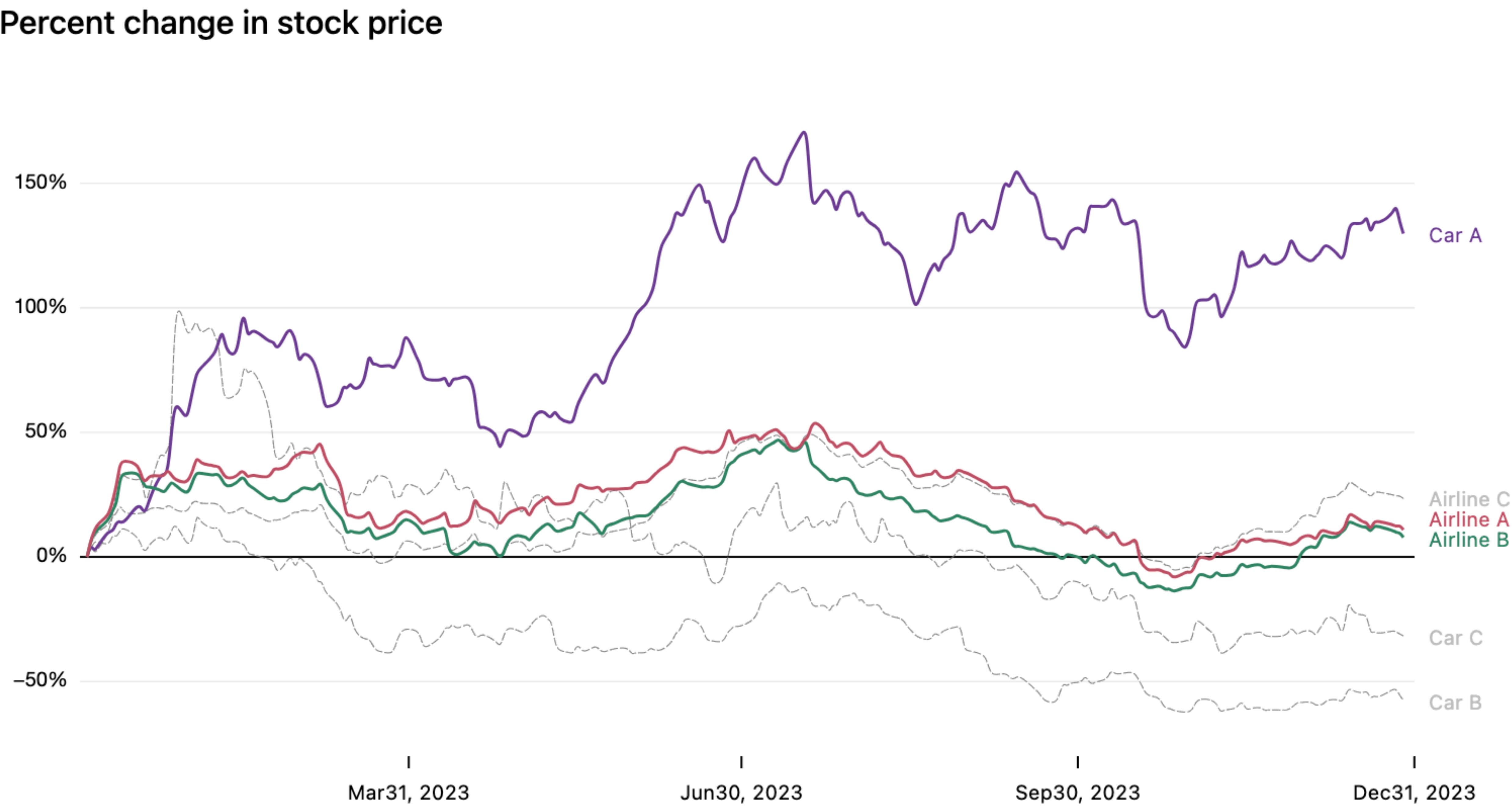
- ☒ Airline A
- ☒ Airline B
- ☐ Airline C
- ☒ Car A
- ☐ Car B
- ☐ Car C
- ☐ Gas A
- ☐ Gas B
- ☐ Gas C
- ☐ Tech A
- ☐ Tech B
- ☐ Tech C

Percent change in stock price



ENFORCING CONTEXT – SUPERIMPOSITION / PRIMARY DATA

- ☒ Airline A
- ☒ Airline B
- ☐ Airline C
- ☒ Car A
- ☐ Car B
- ☐ Car C
- ☐ Gas A
- ☐ Gas B
- ☐ Gas C
- ☐ Tech A
- ☐ Tech B
- ☐ Tech C

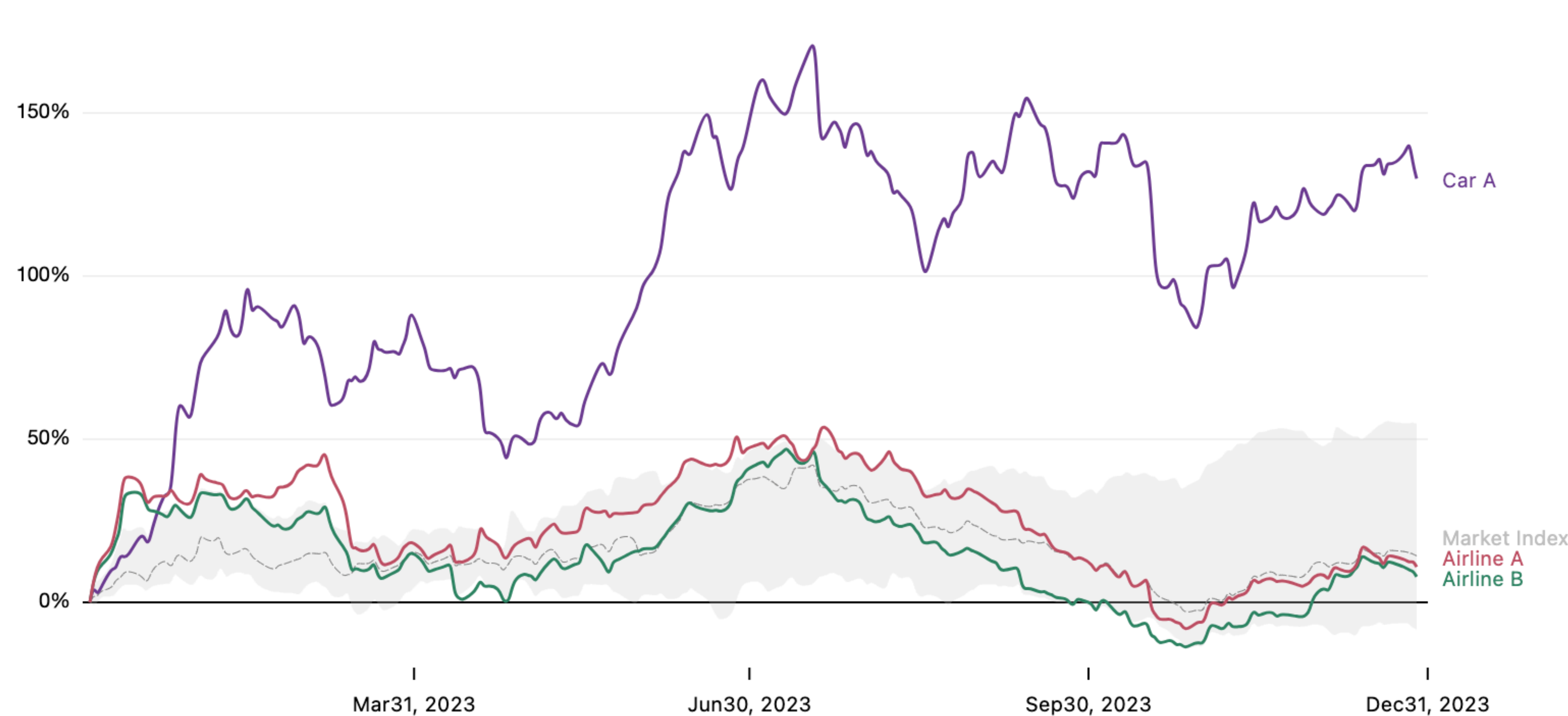


SHOWING STATISTICAL CONTEXT – SUPERIMPOSITION / SUMMARY

- ☒ Airline A
- ☒ Airline B
- ☐ Airline C
- ☒ Car A
- ☐ Car B
- ☐ Car C
- ☐ Gas A
- ☐ Gas B
- ☐ Gas C
- ☐ Tech A
- ☐ Tech B
- ☐ Tech C

Percent change in stock price

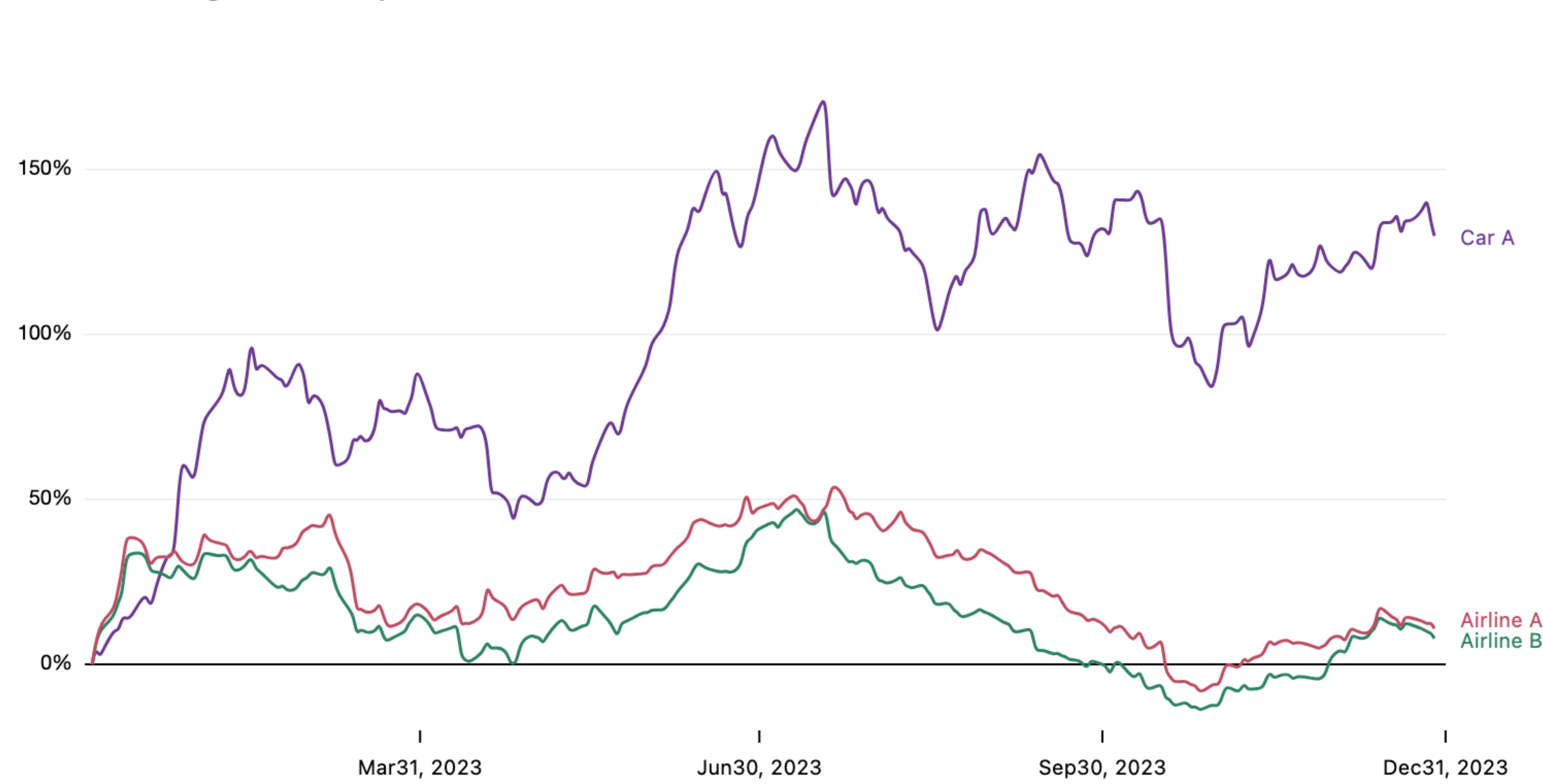
Shaded area represents the middle 50% of all values.



CONTEXT OUTSIDE – JUXTAPOSITION / PRIMARY DATA

- ☒ Airline A
- ☒ Airline B
- ☐ Airline C
- ☒ Car A
- ☐ Car B
- ☐ Car C
- ☐ Gas A
- ☐ Gas B
- ☐ Gas C
- ☐ Tech A
- ☐ Tech B
- ☐ Tech C

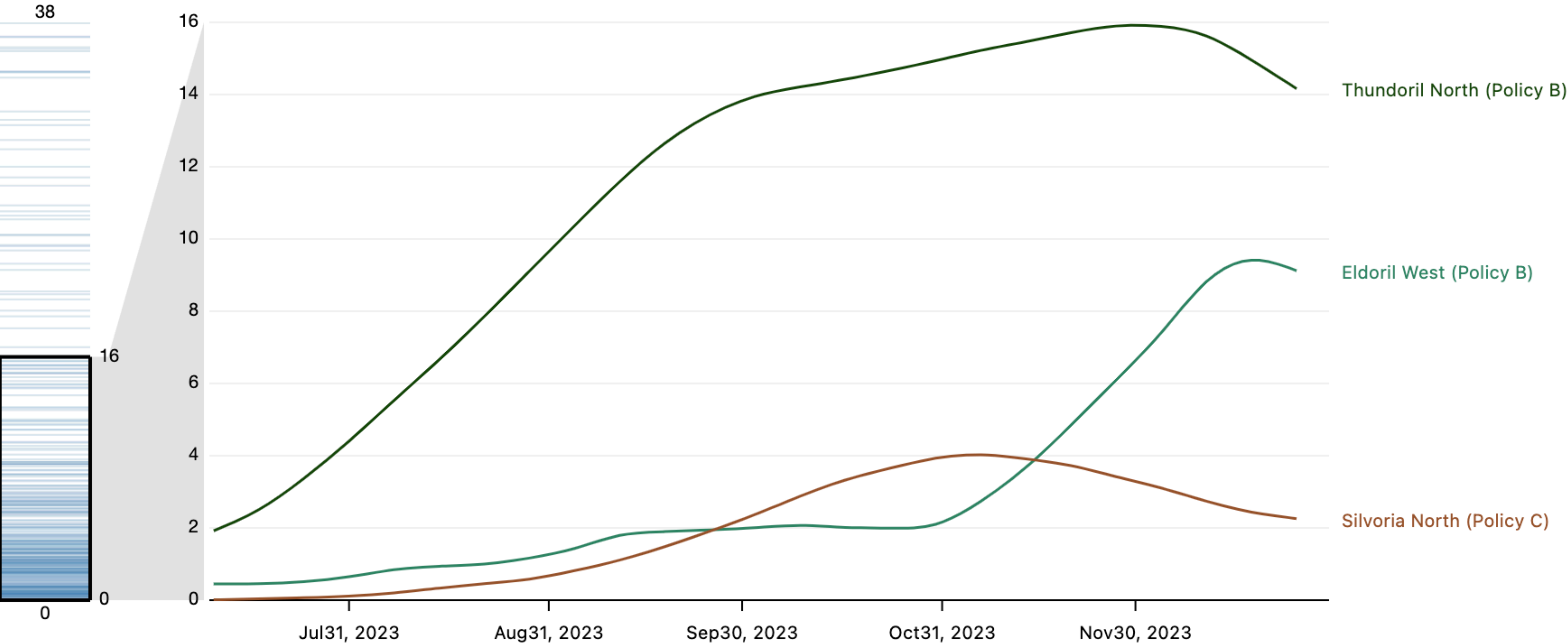
Percent change in stock price



SUMMARY CONTEXT – JUXTAPOSITION / SUMMARY

Infections per million people

Bar on the left highlights the range of selection among all data.



EVALUATION

Red-Team Study (Production):

Challenge participants to deceive w
and w/o guardrails

Validation Study:

Measure how deceptive charts are

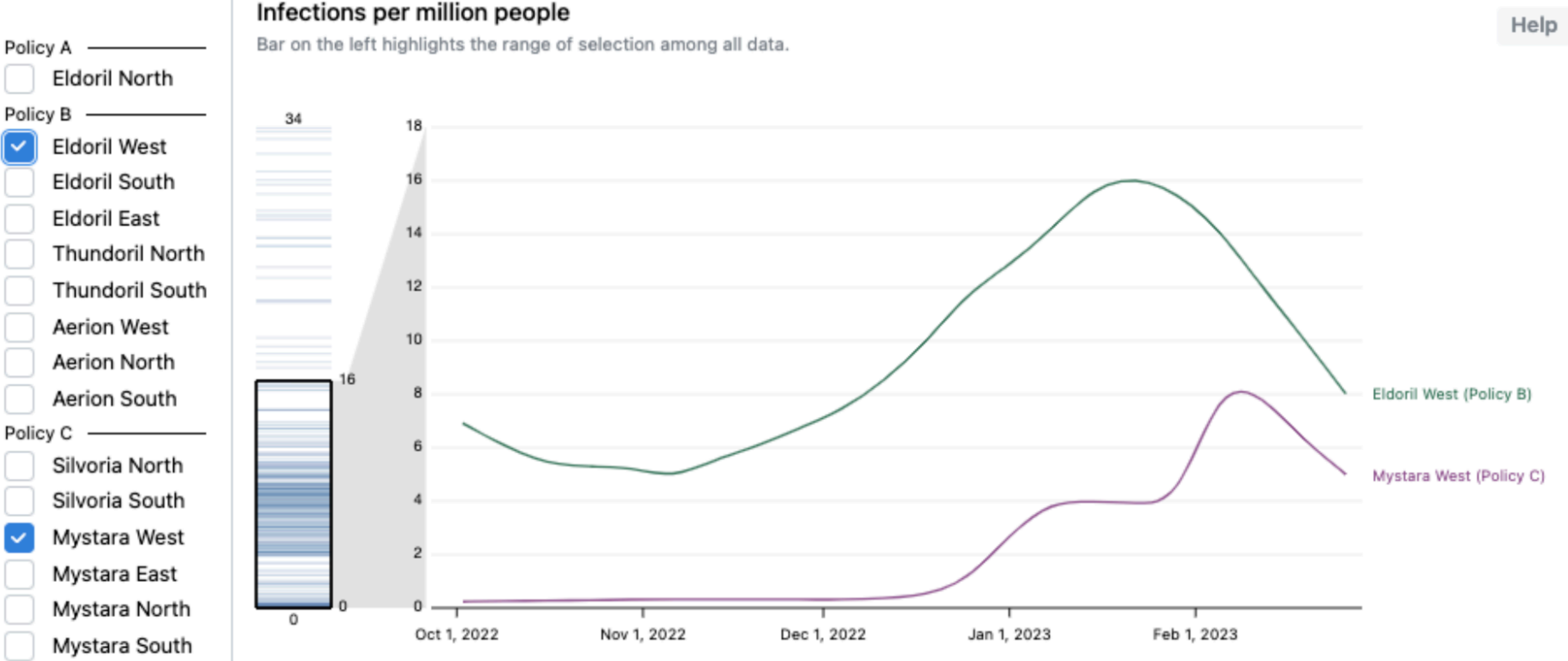
RED-TEAM SCENARIOS

Viral Disease & Stocks

Viral Scenario:

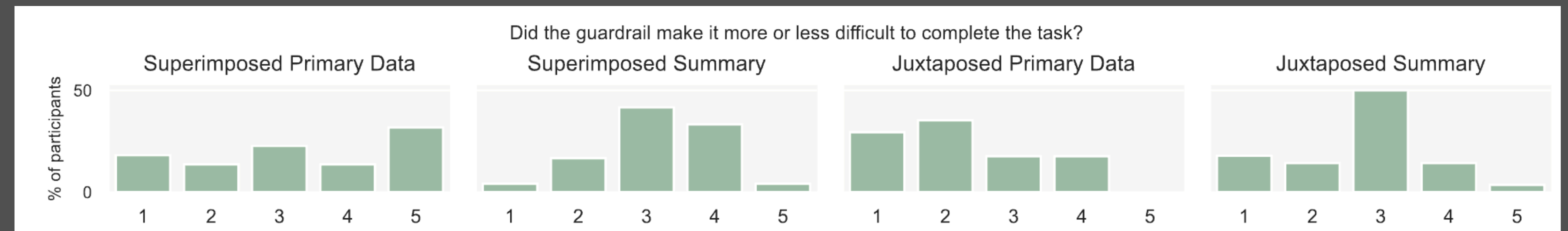
imagine that you work for a public health agency and are in charge of promoting a health policy campaign against a viral disease that, judging by the data alone, does not appear to be the most effective policy.

STIMULUS



Add a short text caption that will go along with your visualization: *

RESULTS



Juxtaposed Primary is a **Cherry-Picking Machine**

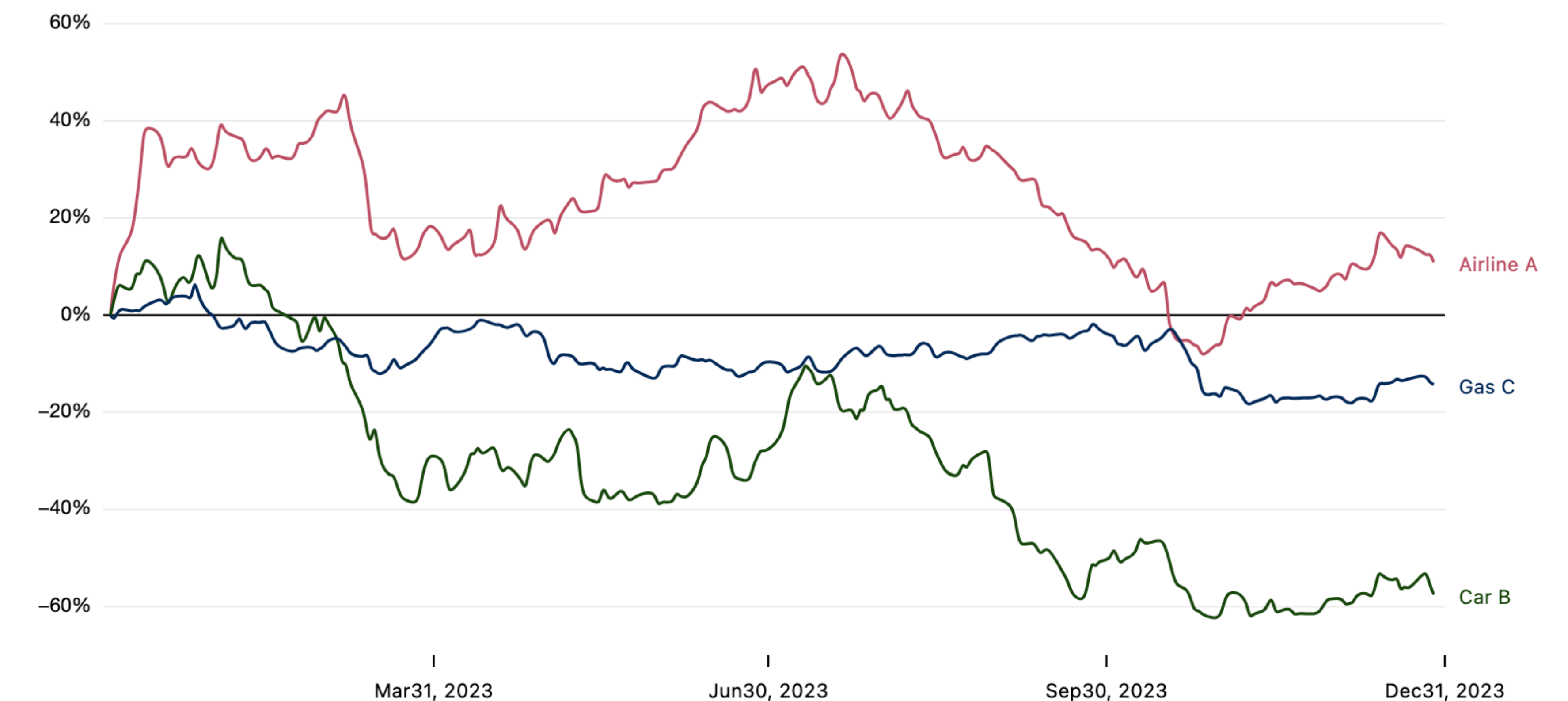
Superimposed Primary makes it **harder to cherry-pick**

Others show little effect

didn't get it?

“ All stocks are not doing well this year, but Airline A is the only one that actually returns some gains, whereas Gas and Car prices have dropped.

Percent change in stock price



VALIDATION STUDY

Do Guardrails Work?

“Investment” task: make monetary decision based on information

You will need to travel to **Eldoril North (Policy A)** for work.

You've come across this visualization and the accompanying caption.

Please review the visualization and the caption, and then answer the questions below **based solely on this information**.

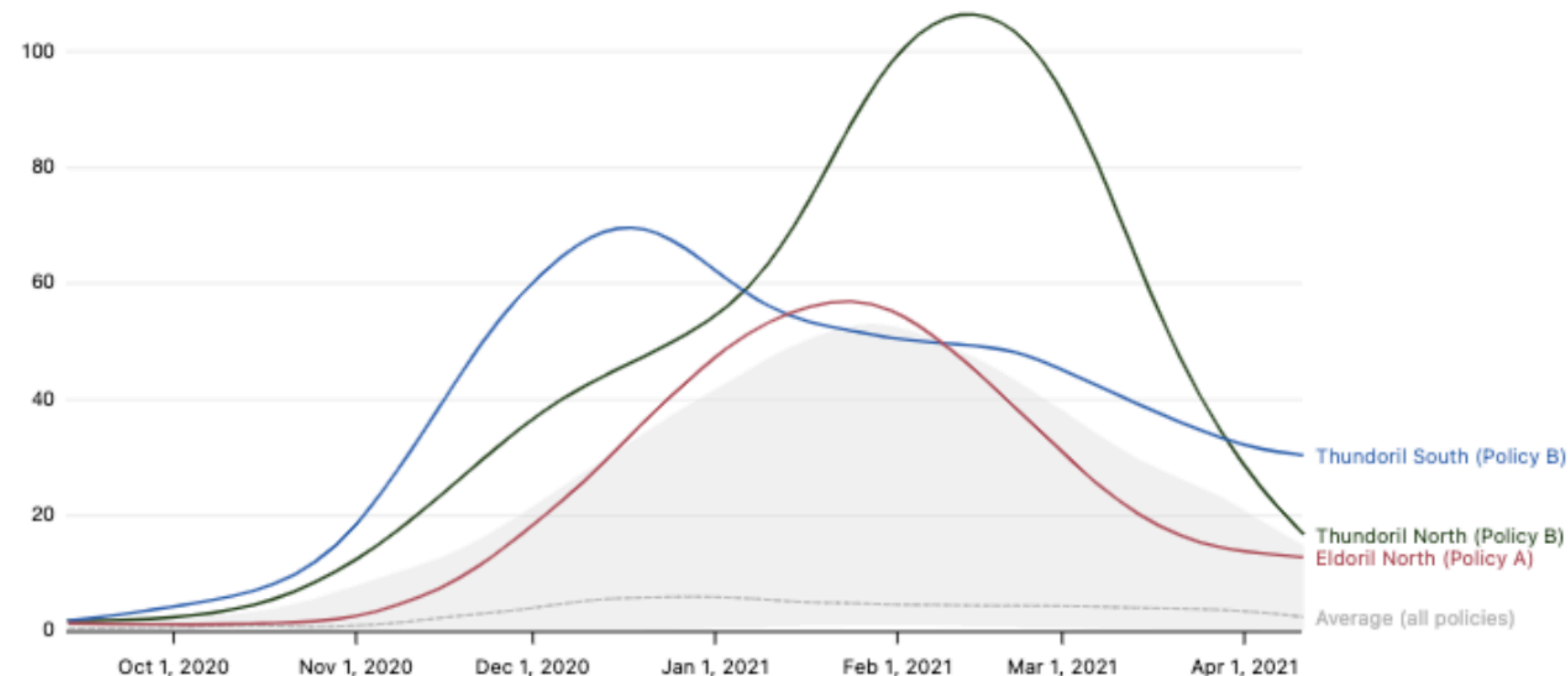
(Please try to not rely on other visualizations you've seen.)

“ Policy A (red) is the superb policy to manage a sudden boom in infections

Infections per million people

Shaded area represents the middle 50% of all values.

Help



Based on this information, before traveling to **Eldoril North (Policy A)** I would buy this much insurance:

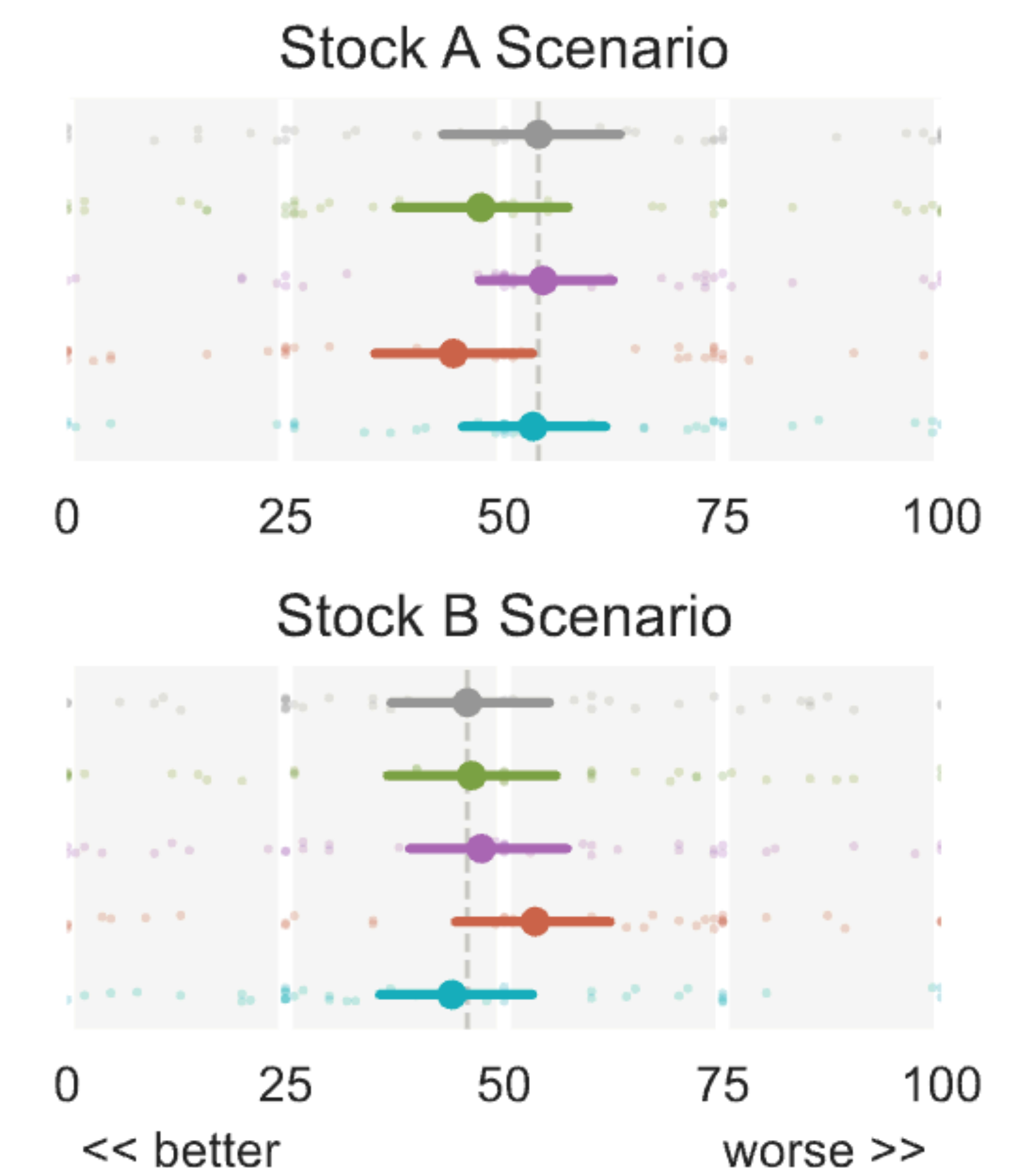
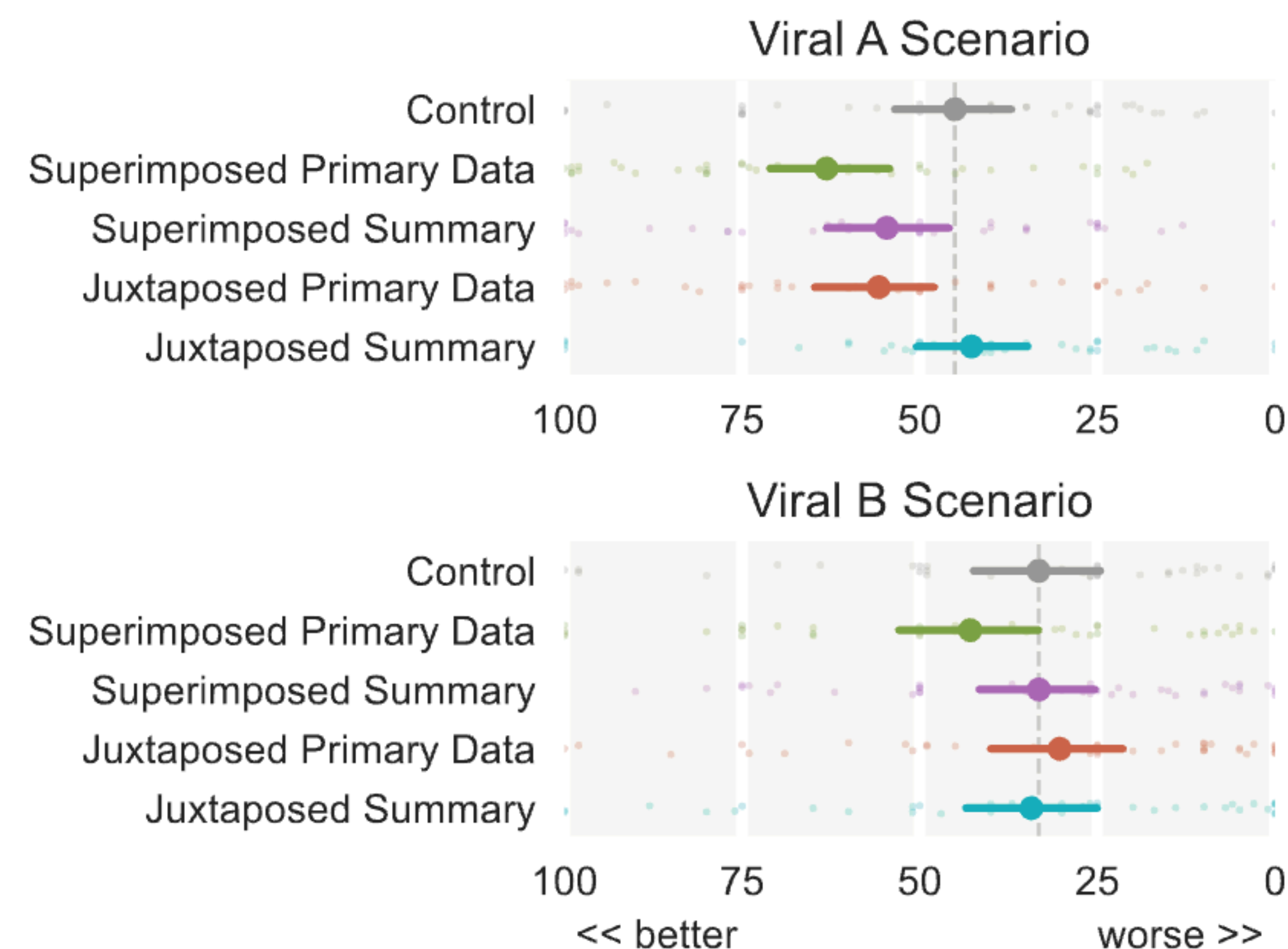
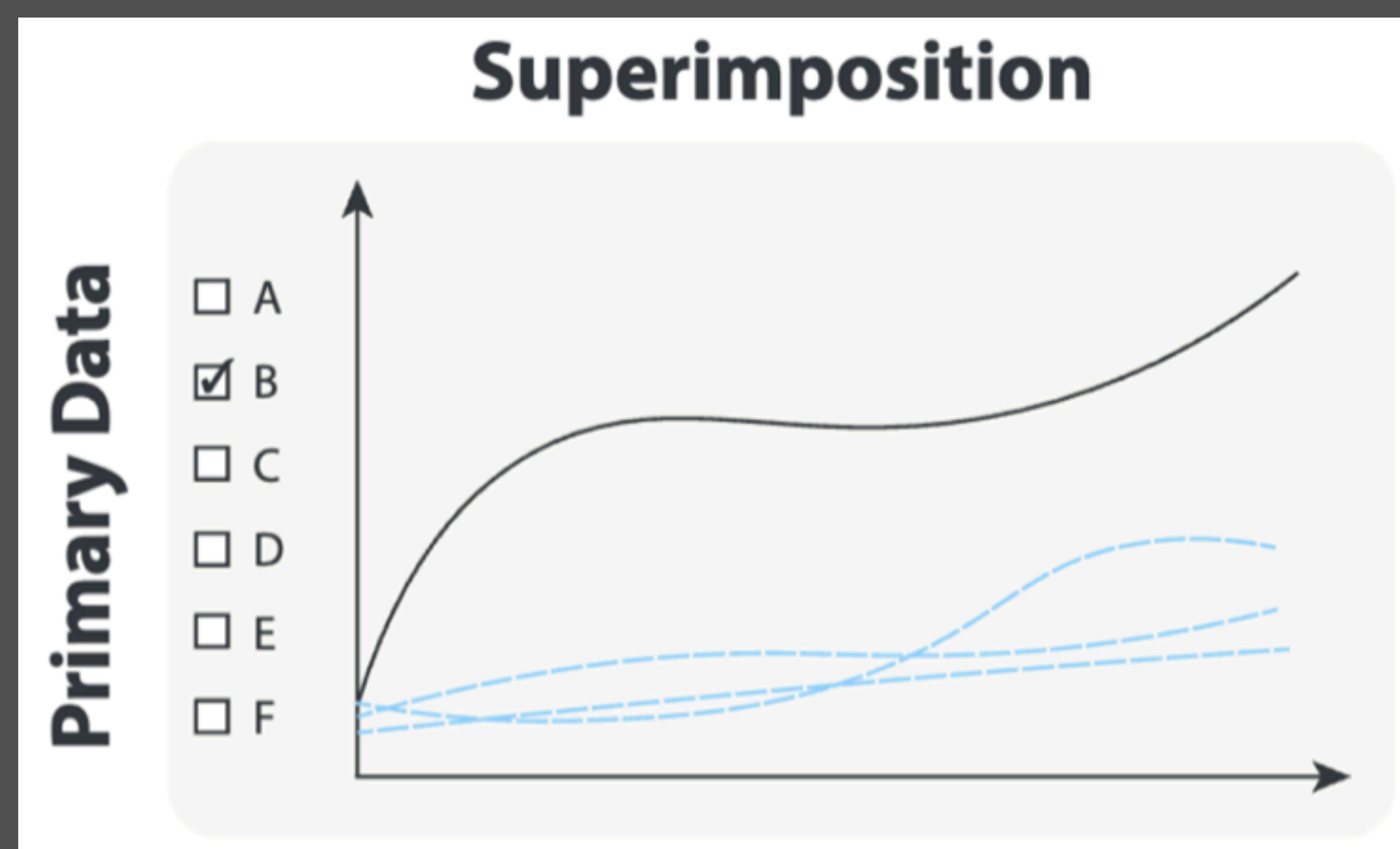
(\$0 = no risk of getting sick, \$100 = very high risk of getting sick)



RESULTS

Differences by Scenario: **More effective in "extreme" scenarios**

Superimposed Primary Data most promising
Juxtaposed Summary no better than baseline



DISCUSSION

Superimposed Primary seems like a promising design against cherry-picking

easy to implement

easy to understand

Juxtaposed Primary design useful as a way to make data explorers more usable

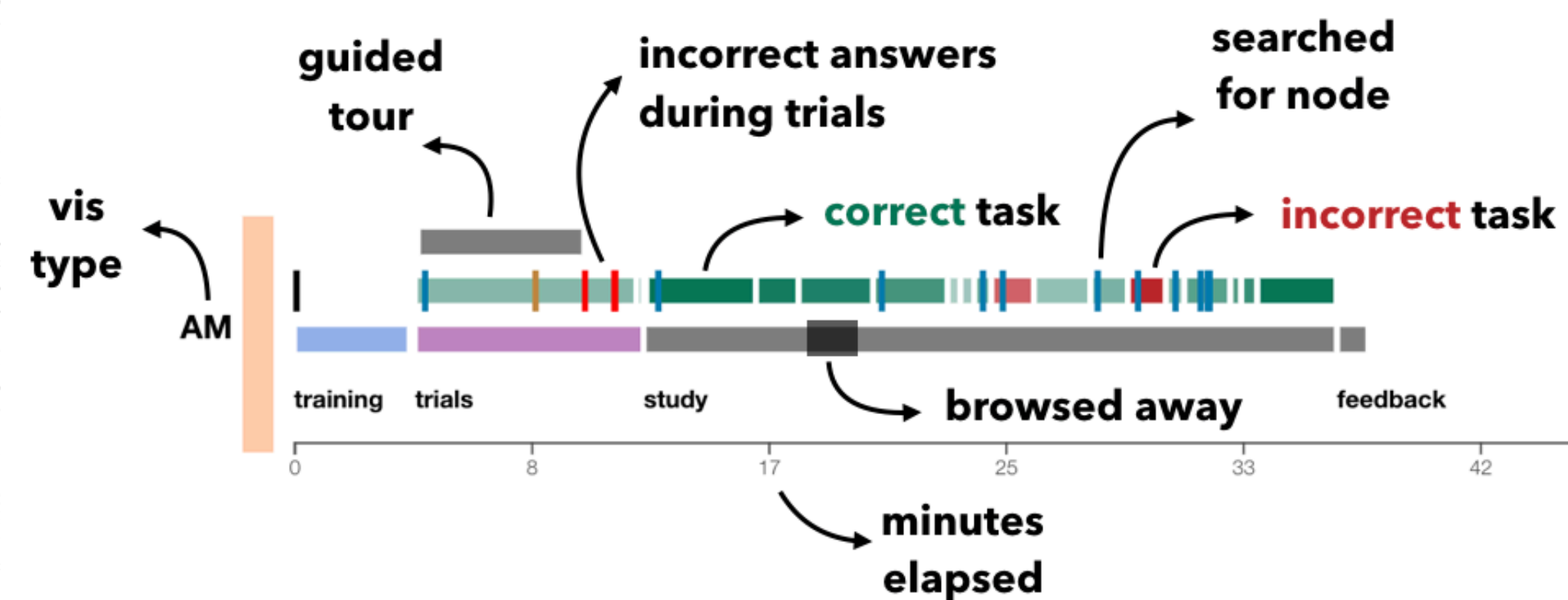
EMPIRICAL & THEORETICAL WORK

Evaluating Complex Systems

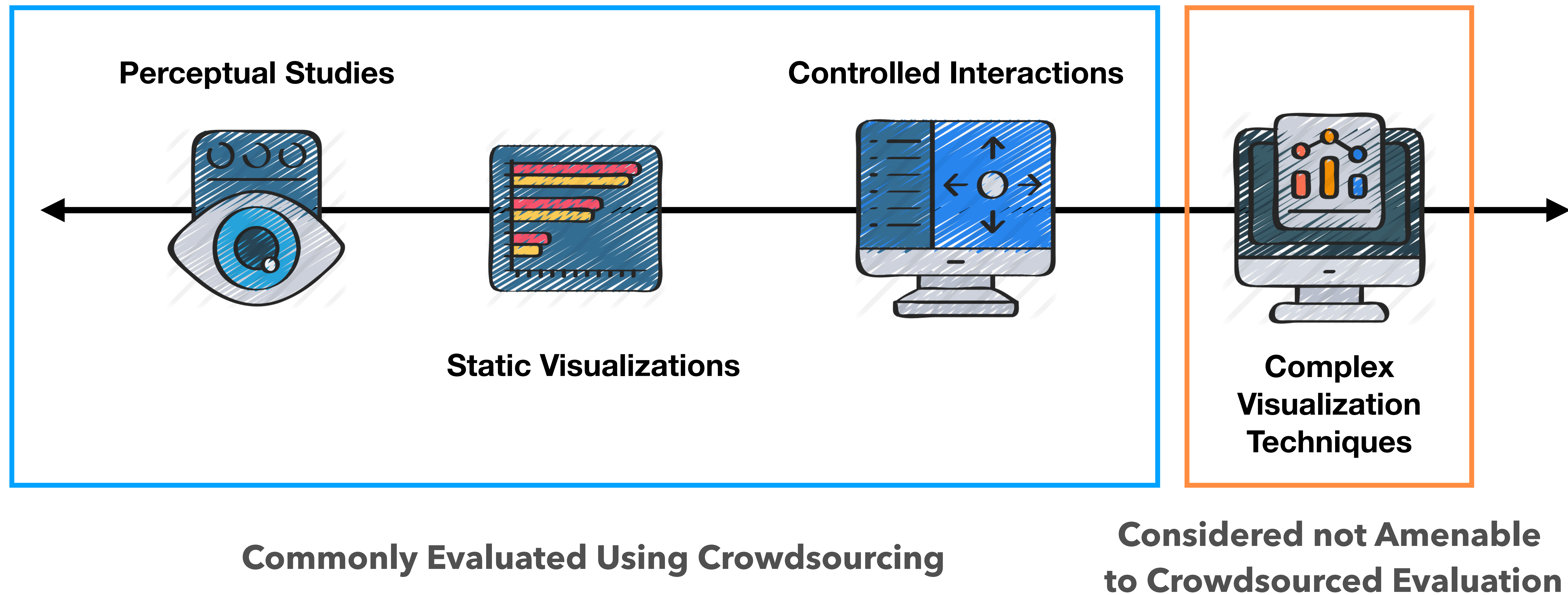


Zach Cutler, Jack Wilburn, Carolina Nobre, Lane Harrison et al.

reVISit: Empirically Evaluating Complex Interactive Visualization Techniques



[CHI 2020, CHI 2021 & VIS 2023]



CAN WE DO QUANTITATIVE EVALUATION WITH COMPLEX SYSTEMS?

Yes We Can!

- Picking the right techniques
- Evidence-based design
- Design validation
- Careful training
- Good compensation
- Interesting Tasks

Pushing the boundary of what can be
evaluated using crowdsourcing

**YET: COMPLEX STUDIES
ARE HARD TO SET UP**

PERILS OF ONLINE VISUALIZATION EXPERIMENTS

There is no **end-to-end platform** for building **interactive experiments**

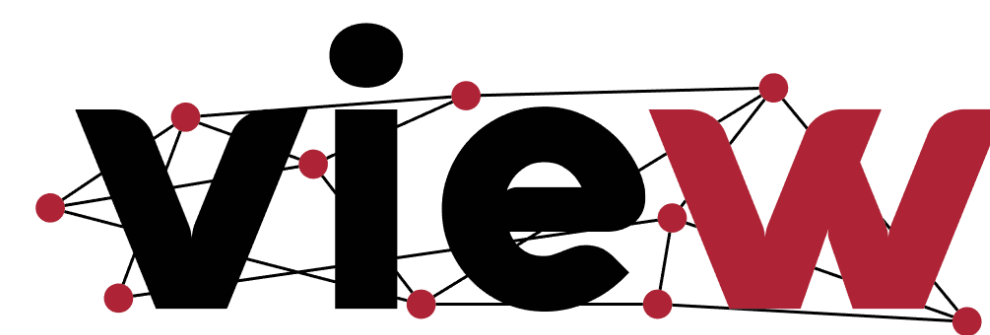
The support for **types of stimuli** in survey tools are limited

Lack of experiment **debugging** features

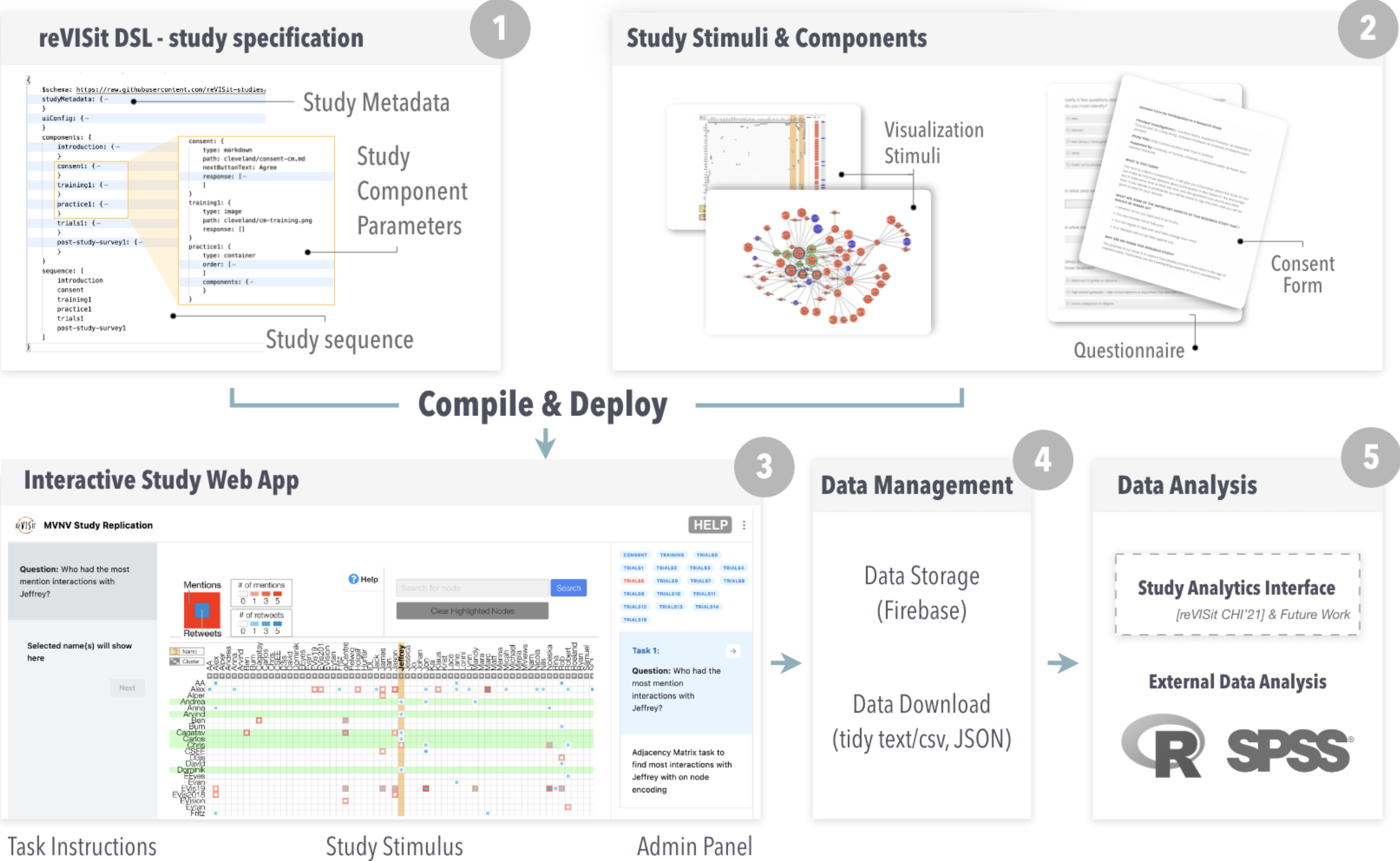
No built-in **data collection** and **provenance tracking** functions



A TOOLKIT FOR DATA
VISUALIZATION EXPERIMENTS



The Scope of reVISit



4

Data Management

Data Storage (Firebase)

Data Download (tidy text/csv, JSON)

5

Data Analysis

Study Analytics Interface
[reVISit CHI'21] & Future Work

External Data Analysis

R SPSS

WHAT IS REVISIT: INFRASTRUCTURE FOR VIS STUDIES

Set up studies with all components (consent, training, trials, tasks, surveys, etc.) including sophisticated **study designs** (randomization)

Simple data tracking and data export

Simple to deploy and run studies

Fully reproducible and open. You can share your whole study setup for anyone to re-run, without having to have a license for software (qualtrics).

REVISIT SPEC

Components & Inheritance

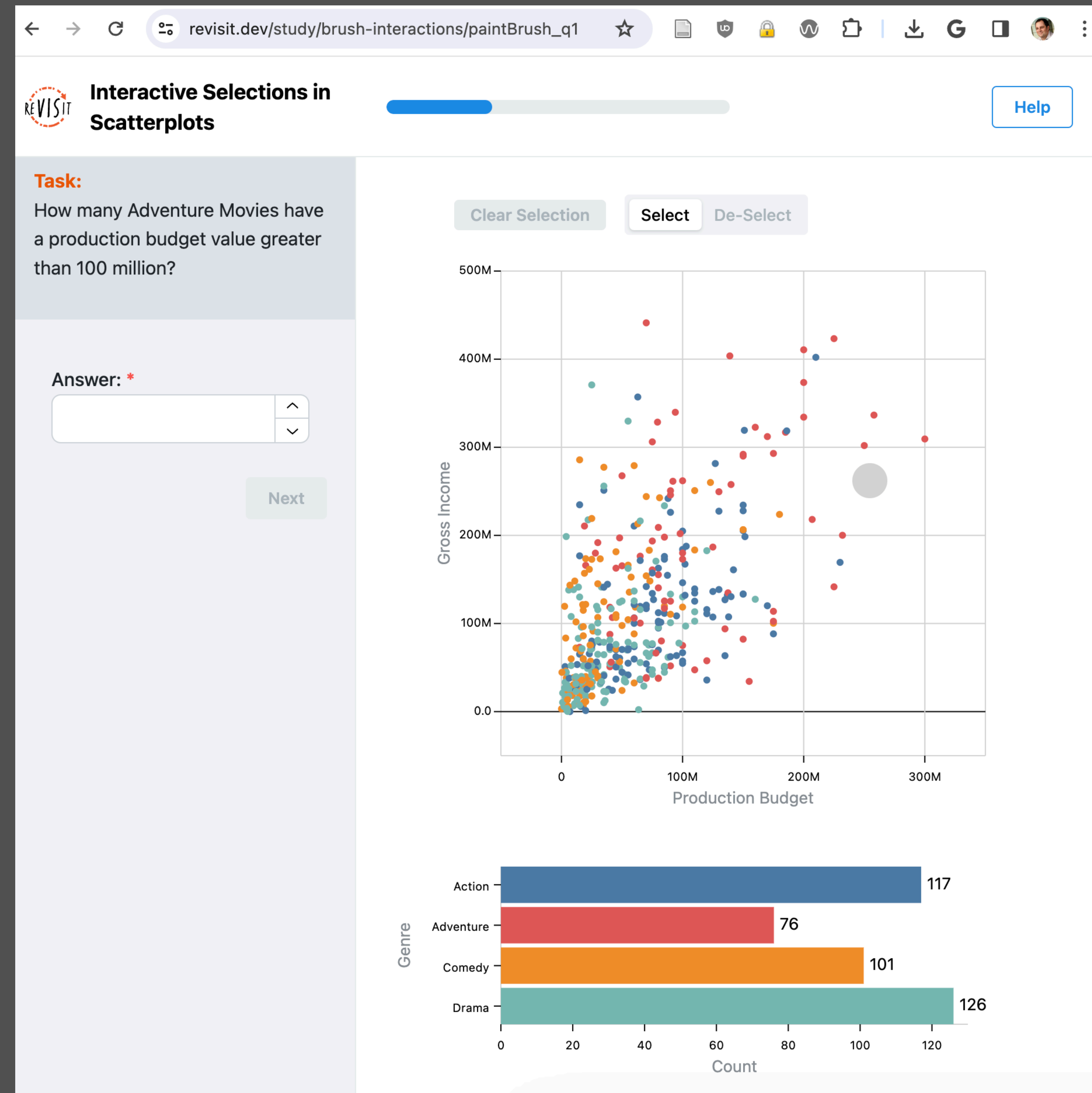
```
"baseComponents": {
  "bar-chart": {
    "type": "website",
    "response": [
      {
        "id": "barChart",
        "prompt": "Your selected answer:",
        "required": true,
        "location": "belowStimulus",
        "type": "iframe"
      }
    ],
    "path": "basic-questionnaire-study/assets/bar-chart.html",
    "instructionLocation": "aboveStimulus"
  }
}

"bar-chart-1": {
  "baseComponent": "bar-chart",
  "description": "A trial for the user to click the largest bar",
  "instruction": "Click on the largest bar",
  "parameters": {
    "barData": [0.32, 0.01, 1.2, 1.3, 0.82, 0.4, 0.3]
  }
},
"bar-chart-2": {
  "baseComponent": "bar-chart",
  "description": "A trial for the user to click the smallest bar",
  "instruction": "Click on the smallest bar",
  "parameters": {
    "barData": [1.2, 1.2, 1.2, 1.3, 0.82, 0.4, 0.3]
  }
}
```


REVISIT SPEC

Study Design

```
"sequence": {  
  "order": "fixed",  
  "components": [  
    "introduction",  
    "consent",  
    "tutorial",  
    {  
      "order": "random",  
      "components": [  
        "paintBrush_q1",  
        "rectangleBrush_q1",  
        "axisBrush_q1",  
        "sliderBrush_q1"  
      ]  
    },  
    "post-study-survey",  
    "survey"  
  ]  
}
```



<https://revisit.dev>

DOCUMENTATION & COMMUNITY

Documentation & Tutorials on Website
Community participation via slack etc.
Tutorial at IEEE VIS 2024



ReVISit


About

Community

Documentation

v1.0.0-beta11 ▾

GitHub 

Demo 



Search



A study creation platform allowing you to quickly create, publish, and disseminate your customized visualization study.

About ReVISit

Try The Demo

Get Started



Flexible And Powerful

ReVISit is designed with researchers from all disciplines in mind. It is simple enough



Focus on What Matters

ReVISit allows researchers to focus on the visual stimuli without the hassle of setting



In Depth Analysis

With the Analysis Dashboard, you can investigate the results from your study with

CAN WE DO QUALITATIVE EVALUATION ONLINE?

Maybe?

- Think Aloud & Provenance

The screenshot displays a qualitative evaluation interface. At the top, there is an audio waveform and a progress bar with three colored dots (orange, orange, blue). Below the audio player, there are controls for 'Play' and 'Pause', and a timestamp '3/18/2024, 3:07:31 PM'. The main area is labeled 'Transcript' and contains a list of text segments, each preceded by a vertical bar. To the right of the transcript, there is a section labeled '+ Tags' with a list of tags: 'Reading question', 'Reading question', 'Stating intent', and several empty tags.

Play Pause 3/18/2024, 3:07:31 PM

Transcript

So for this task, uh, I just need to consider penguins that weigh between 4 and a half and 5.

Okay kilograms, but all that I'll flip a length between 2 and 215.

All right. I'm going to have to be quite precise so I can

I'm I'm confident about the 4 to 5K but 2 to 15.

Okay, so I'm just going to click on select.

All right, so actually 200 to 20.

So I'm just going to have to have to estimate where.

2 15 is

which is approximately there about 75% of that.

Let me just clear that selection. Let me just try this again.

+ Tags

Reading question ×

Reading question ×

Stating intent ×

Pushing the boundary of what can be
evaluated using crowdsourcing



5e5521580ee1b951df544c3c



paintBrush_q4



Play



Pause

3/18/2024, 3:08:05 PM

Transcript

+ Tags

So for this task, uh, I just need to consider penguins that weigh between 4 and a half and 5.

Reading question ✕



Okay kilograms, but all that I'll flip a length between 2 and 215.

Reading question ✕



All right. I'm going to have to be quite precise so I can

I'm I'm confident about the 4 to 5K but 2 to 15.

Okay, so I'm just going to click on select.

Stating Intent ✕



All right, so actually 200 to 20.

So I'm just going to have to have to estimate where.

2 15 is

which is approximately there about 75% of that.

Let me just clear that selection. Let me just try this again.

Okay. Alright, so that's the 4 to 5K body mass. And then the 2002215.

So that's giving me those results.

so the most common penguin There Is the Gen 2

Answering question ✕



and the least common penguin with 5 is the a deli if I've pronounced that correctly.

Okay clicking.



Task:

Consider only penguins that weigh between 4k and 5k grams (g) and have flipper lengths between 200 and 215mm. What is the most and least common type of penguin in this subset?

While answering this question, please verbalize your thoughts, especially any insights you have or problems you run into.

Most common Penguin: *

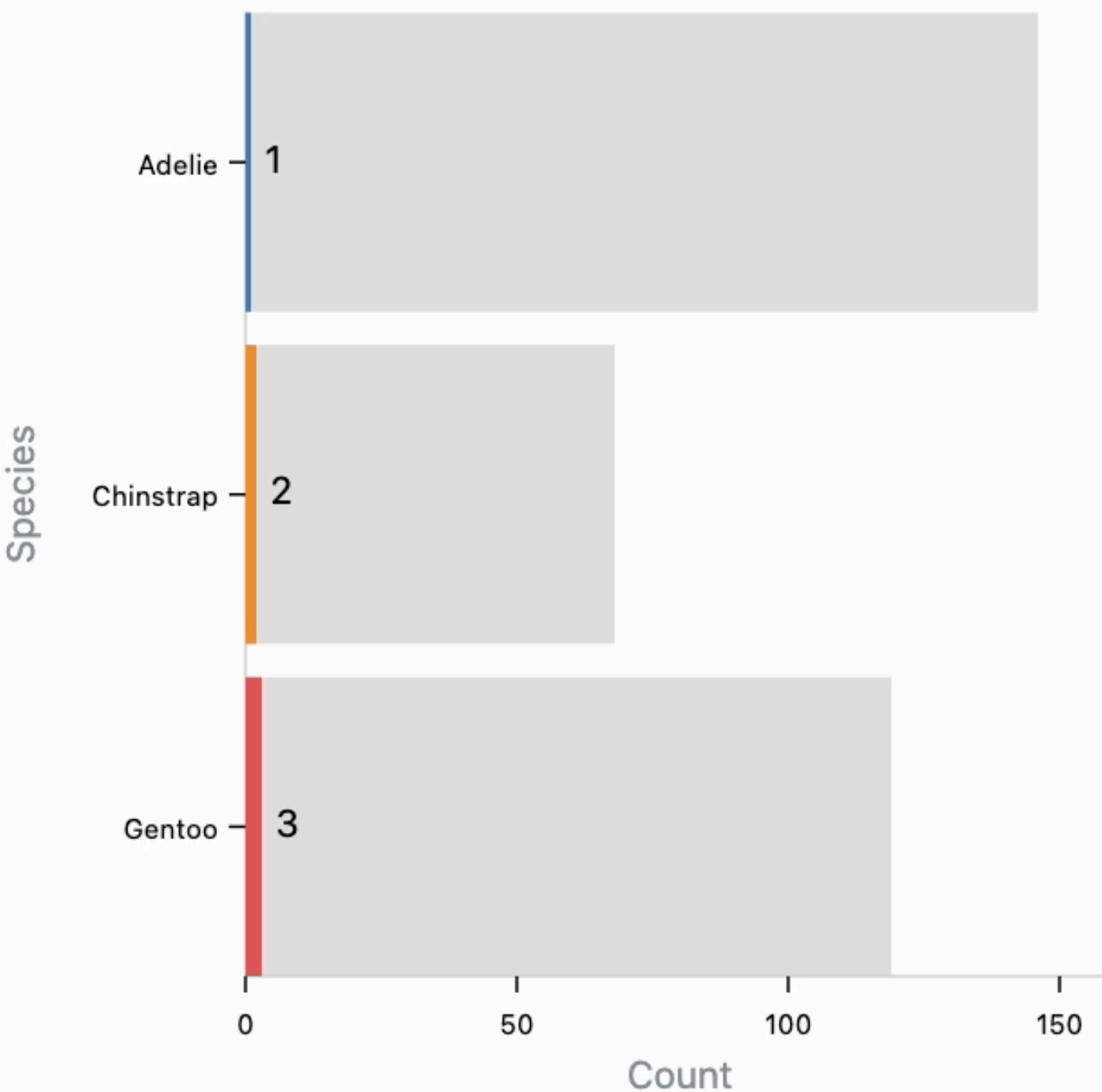
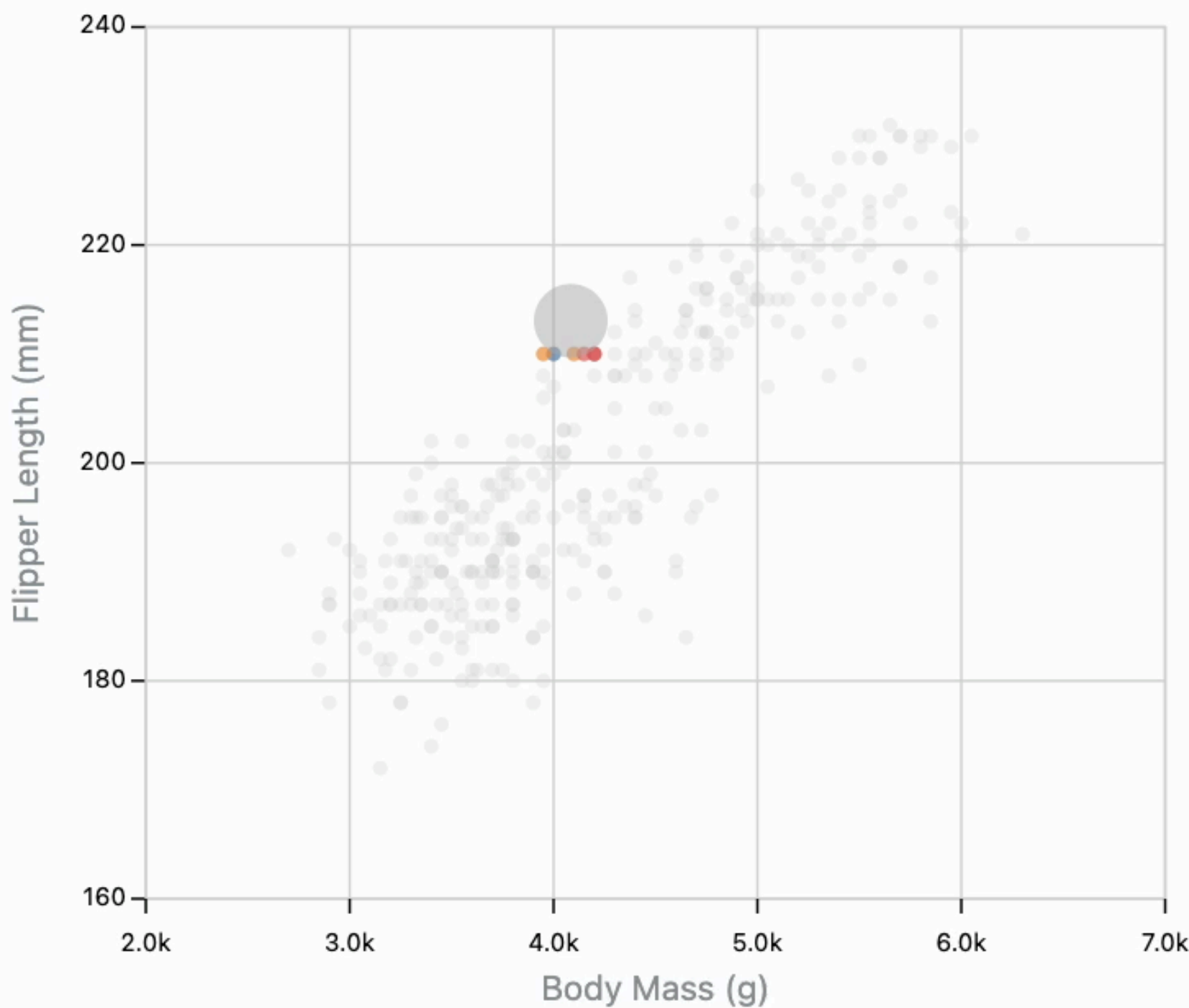
Least common Penguin: *

Next

Clear Selection

Select

De-Select





Play Pause 3/18/2024, 3:07:31 PM

Transcript

So for this task, uh, I just need to consider penguins that weigh between 4 and a half and 5.

Okay kilograms, but all that I'll flip a length between 2 and 215.

All right. I'm going to have to be quite precise so I can

I'm I'm confident about the 4 to 5K but 2 to 15.

Okay, so I'm just going to click on select.

All right, so actually 200 to 20.

So I'm just going to have to have to estimate where.

2 15 is

which is approximately there about 75% of that.

Let me just clear that selection. Let me just try this again.

+ Tags

- Reading question ×
- Reading question ×
- Stating intent ×
-
-
-
-
-
-

FUTURE WORK: BETTER SCAFFOLDING FOR STUDIES

reVISit: Scalable Empirical Evaluation of Interactive Visualizations

Community Input

Collaborators
Core Community

Community Workshops
Broader Community

User Services and Resources

Documentation / Examples Replications Community Engagement

----- ↓ ----- **Inform Software Development** ----- ↓ ----- ----- ↓↑ ----- **Synergistic Activities** ----- ↓↑ -----

Core Infrastructure: Software Components

——— ● New Component —— ● Prototype —— ● Robust ———

Aim 1: Study Infrastructure and Multilevel Instrumentation

Study Scaffolding Integration with Crowdsourcing Platforms

Component Registry Capturing Insights / Rationale

Provenance Tracking Designing Trainings

Aim 2: Data Transformation and Visualization Methods

Data Transformation Event Sequence Visualization

Trials Visualization Study Results Visualization

Study Replay Interface Qualitative Coding Tools

REFLECTIONS

A STEW OF RESEARCH TOPICS

Doing different things is interesting!

Keeps me engaged

Role of Engineering

Building / maintaining / documenting
useful things increases impact

reVISit, Persist, UpSet

Crossover Benefits

Provenance -> reVISit & Persist

reVISit -> Misinformation

Engineering infrastructure: benefits all

Alexander Lex
<http://vdl.sci.utah.edu>



A Hodgepodge of Visualization Research: Provenance, User Studies, Misinformation



visualization
design lab

