

# **SEARCHING FOR MY NEXT CHART**



**MUHAMMAD NAKHAE**

**PYDATA AMSTERDAM 2025**



Work as an ai engineer @ Exact

Love photography and collecting cameras @merrygoroundofphotos

Exploring my creative side these days

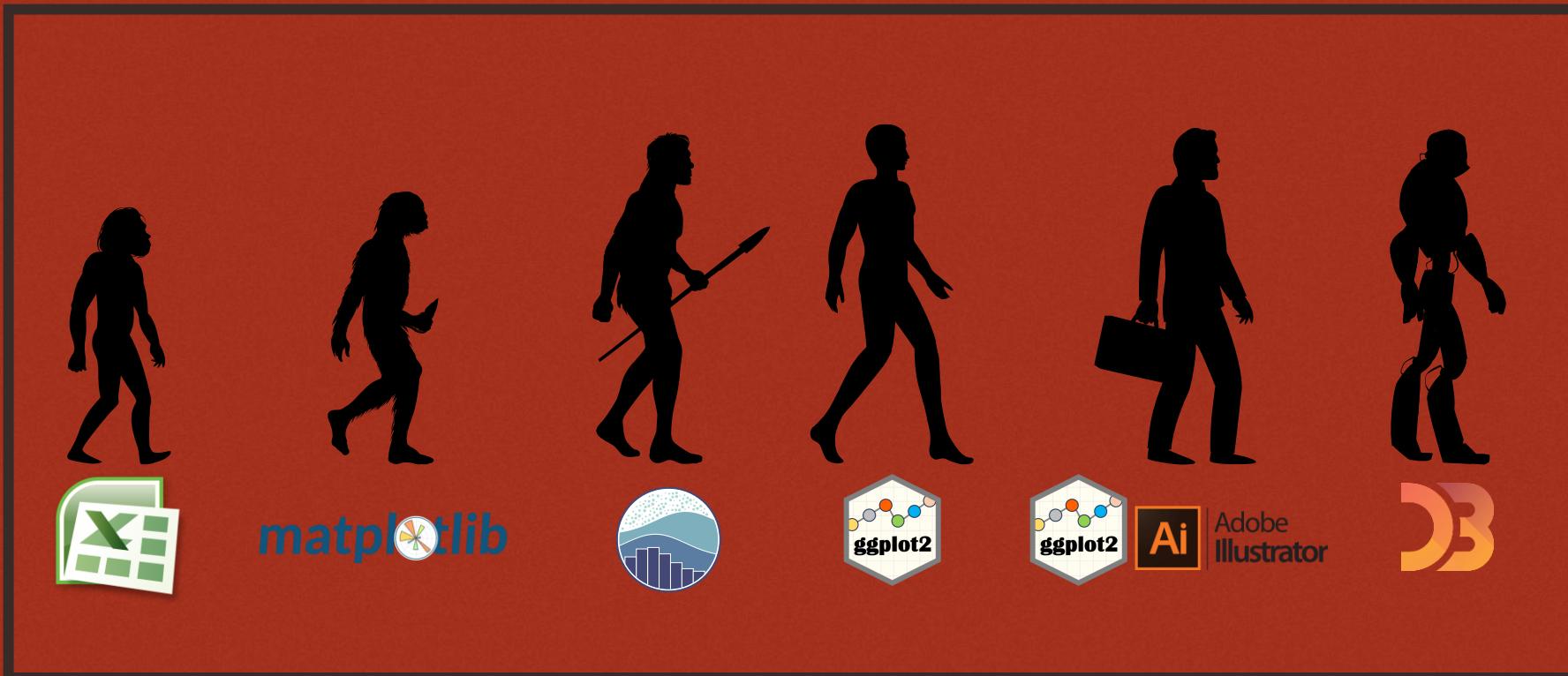
Perfectionist!

# This Talk

is **NOT** a tutorial on how to design a great chart

looks at “data visualization” as a **creative** form, not just a practical tool.

# My “Creative” DataViz Evolution



# My “Creative” DataViz Evolution



# Why ggplot2

Intuitive (based on  
the grammar of  
graphics)

Powerful

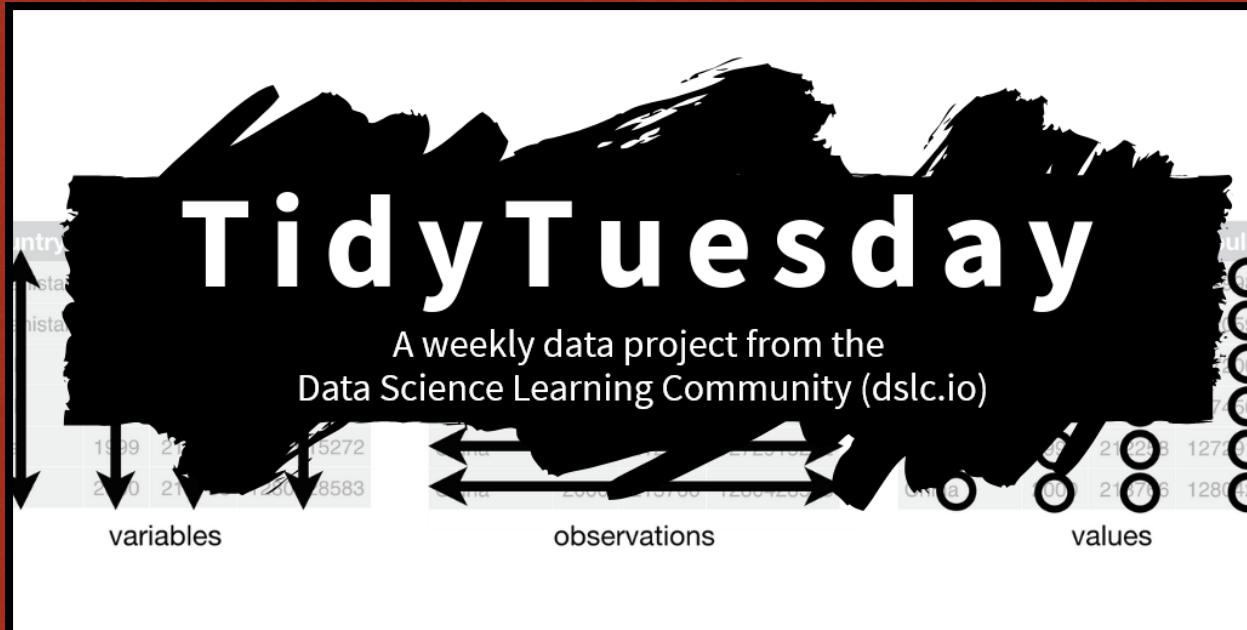
User  
base



Flexible

Developer  
base

# TidyTuesday



## #30DayChartChallenge

April 2022 • 30 Days • 30 Charts • 5 Categories

Comparisons

- 1. part-to-whole
- 2. pictogram
- 3. historical
- 4. flora
- 5. slope
- 6. data day: OWID

Distributions

- 7. physical
- 8. mountains
- 9. statistics
- 10. experimental
- 11. circular
- 12. theme day: The Economist

Relationships

- 13. correlation
- 14. 3-dimensional
- 15. multivariate
- 16. environment
- 17. connections
- 18. data day: OECD

Timeseries

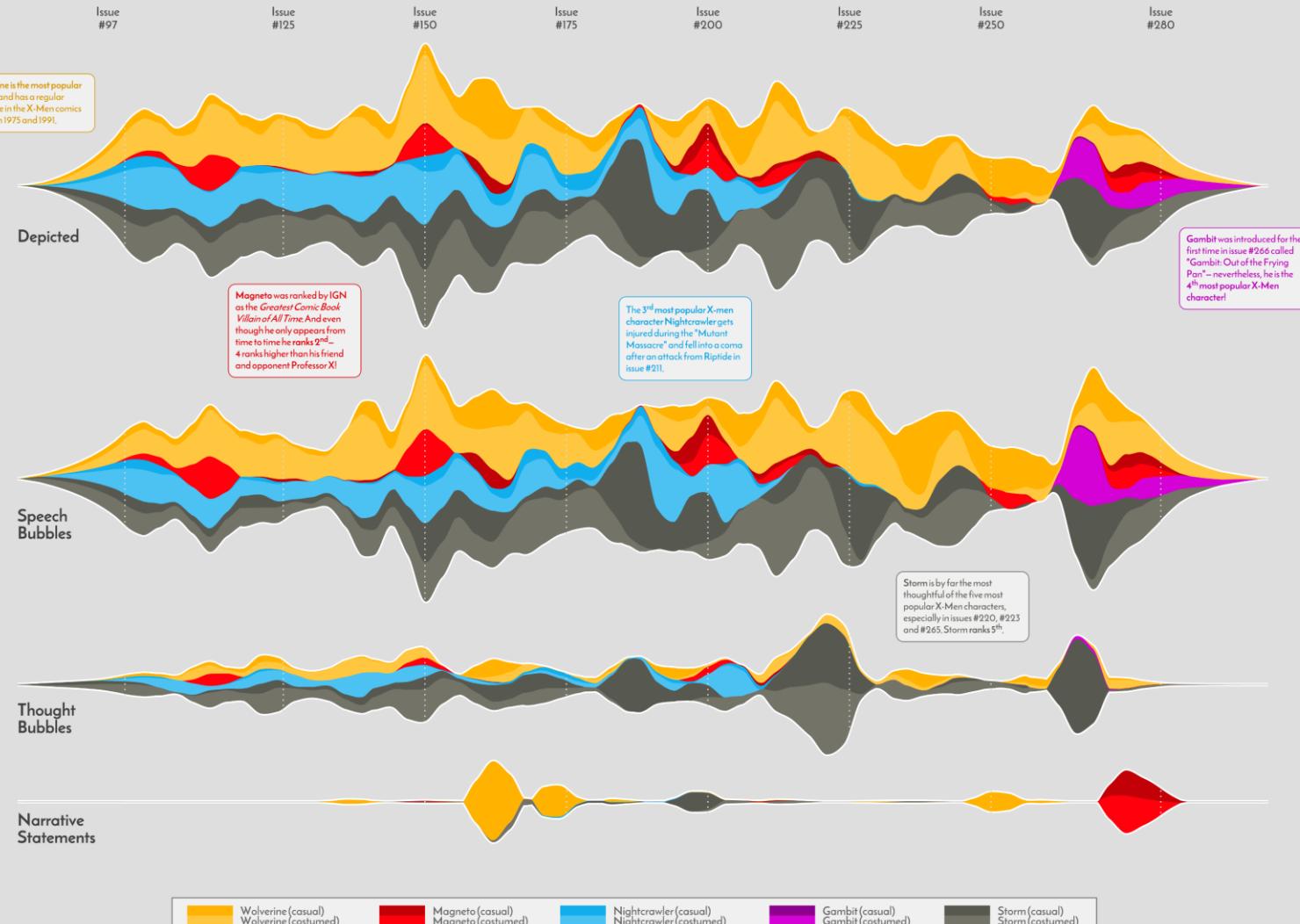
- 19. global change
- 20. new tool
- 21. down/upwards
- 22. animation
- 23. tiles
- 24. theme day: Financial Times

Uncertainties

- 25. trend
- 26. interactive
- 27. future
- 28. deviations
- 29. storytelling
- 30. data day: UN Population

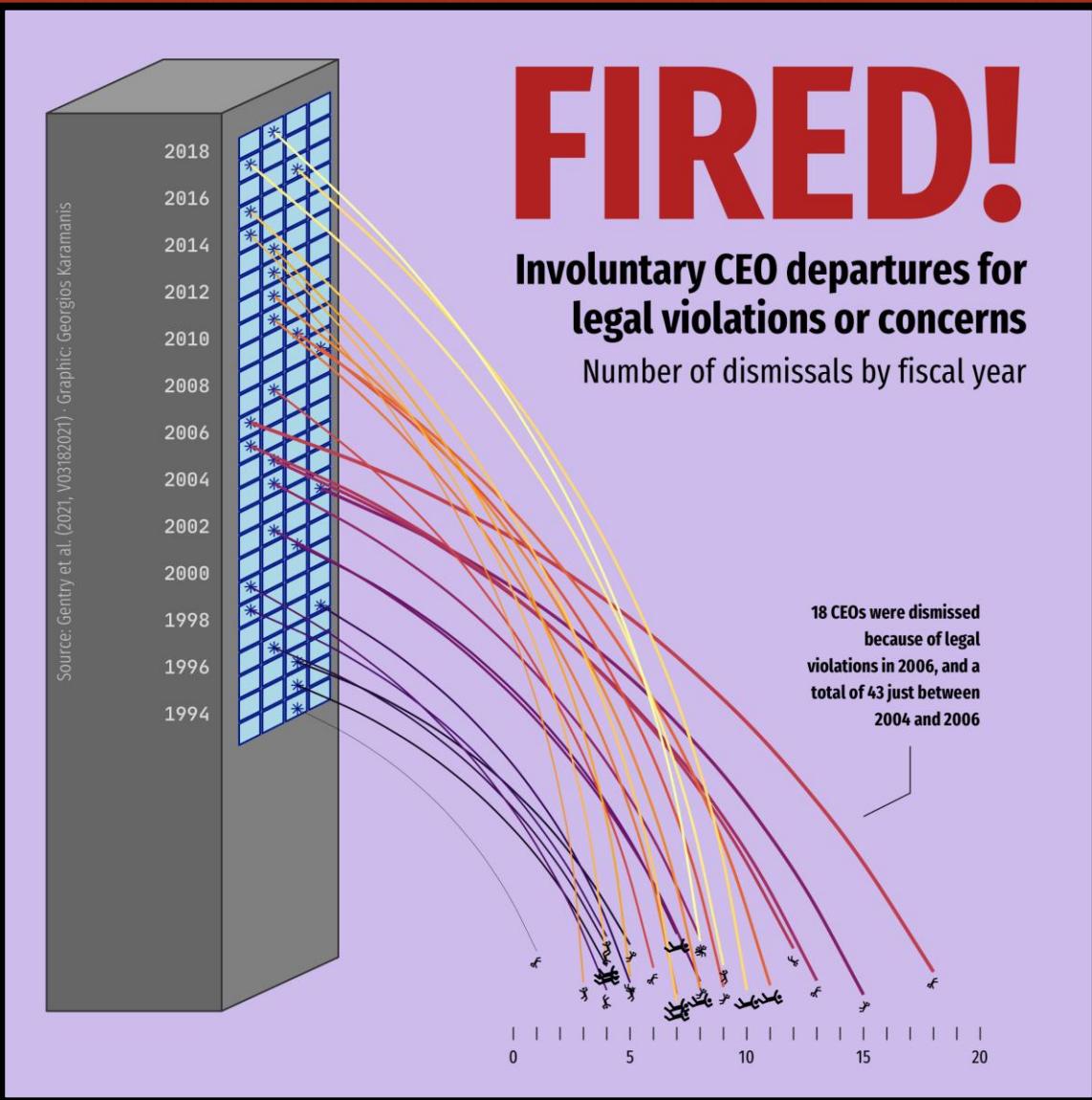
Follow [@30DayChartChall](#) for more!

# Average Appearance of the Five Most Popular X-Men Characters in Chris Claremont's **X-MEN** Comics



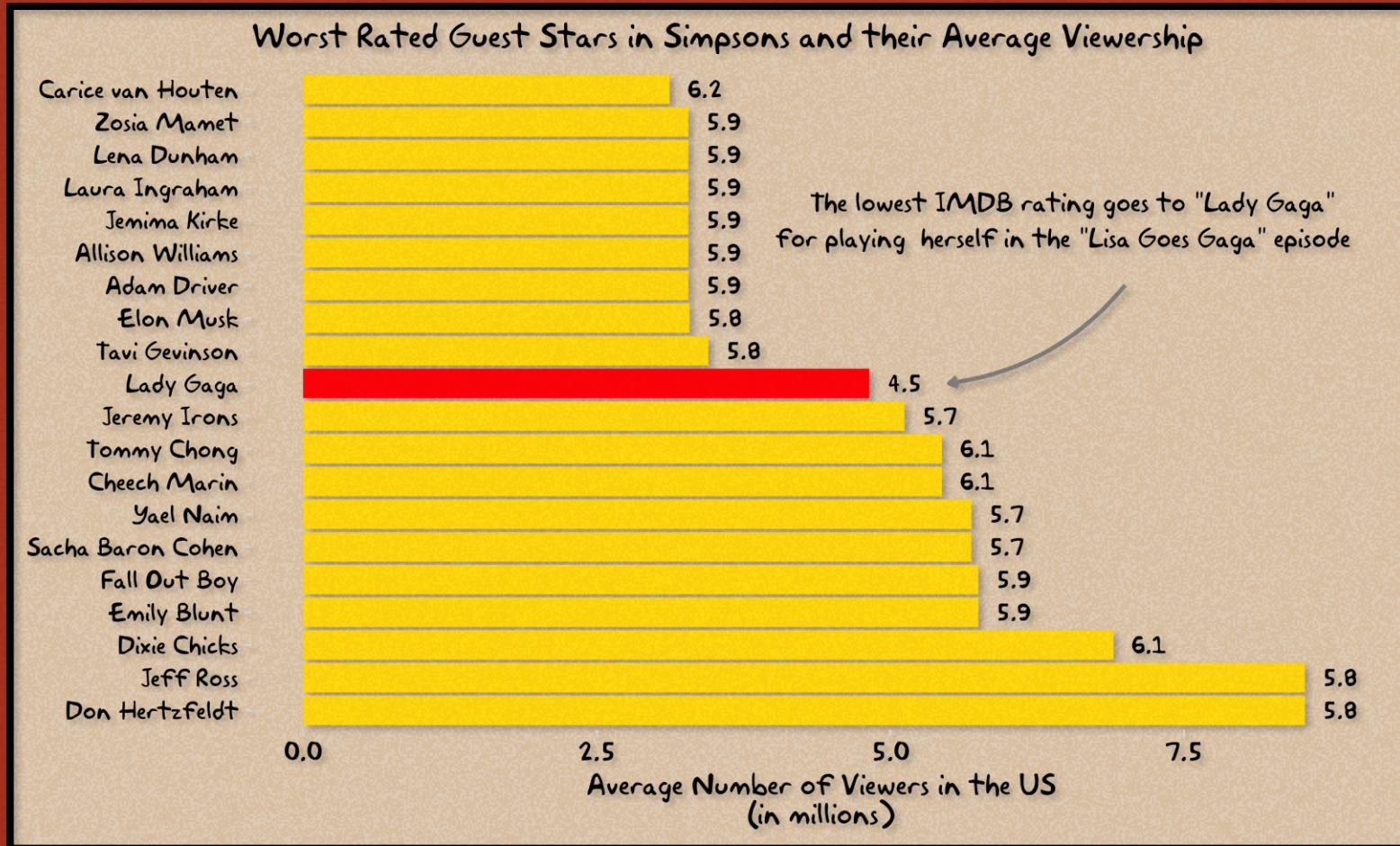
Visualization by Cédric Scherer • Data by Claremont Run Project via Malcolm Barrett • Popularity Scores by ranker.com • Logo by Comicraft

Cédric Scherer Uncanny X-Men by Claremont Run Project

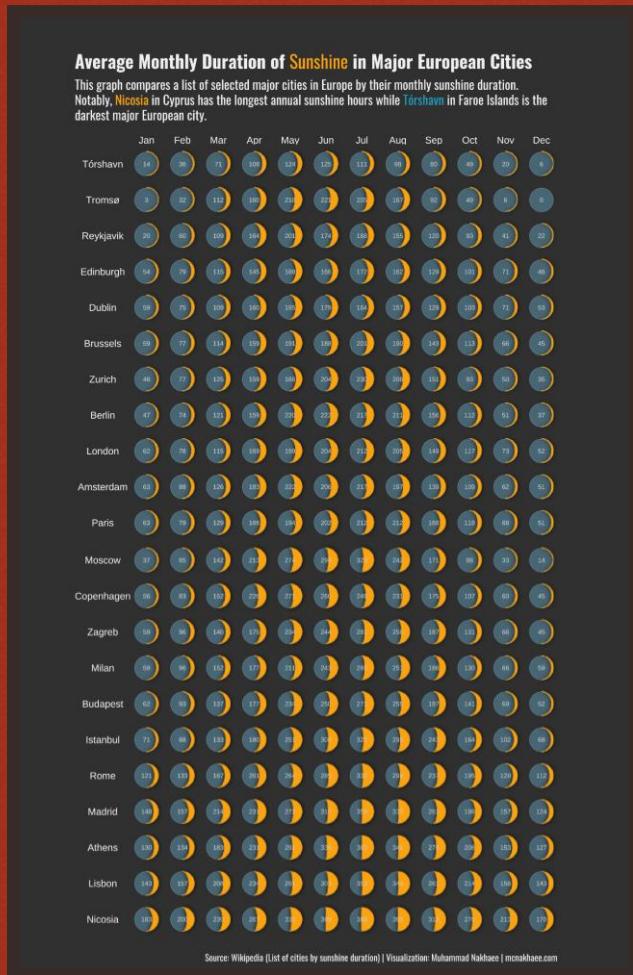


CEO departures, Georgios Karamanis

# My First #TidyTuesday Submission



# #30DayChartChallenge



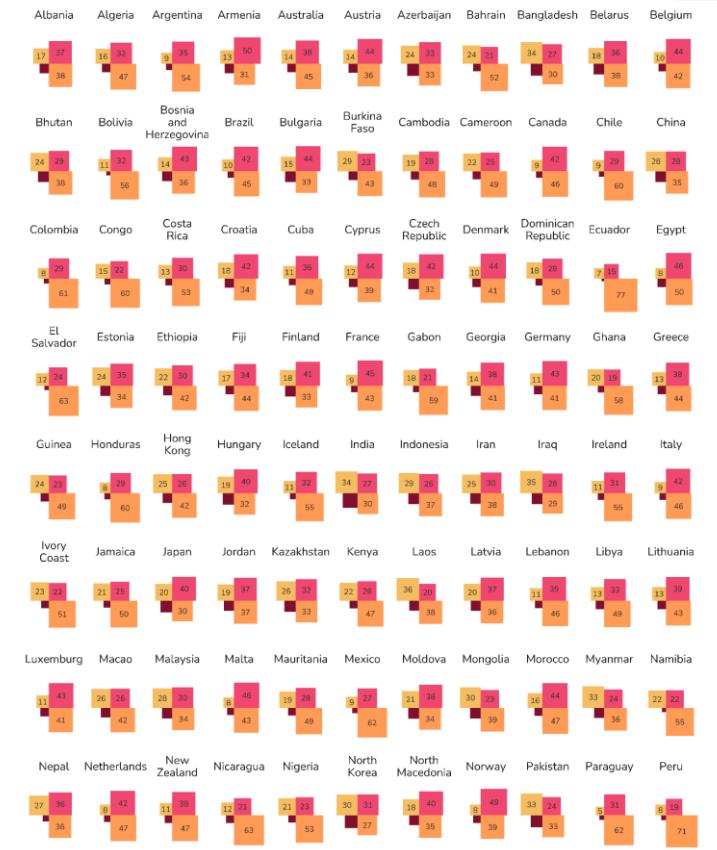
## Rothko's Favorite Mediums

Ink Oil Synthetic polymer paint Watercolor Other



## Blood Type Distribution by Country

Blood type is a classification of blood, based on the presence and absence of particular antibodies and antigens in red blood cells. Blood types are inherited and represent contributions from both parents of an individual. The two most important blood type systems type A, B, AB, and O, with + or - which denotes the RhD status)



# Winning a Kaggle Prize



## Main Missions:

- Analyze and understand data to influence product or business decisions
- Build and/or the data infrastructure that my business uses storing, analyzing, and operationalizing data

7,500-9,999 Median Compensation

Under 1 year Median Coding XP

Under 1 year Median ML XP

## Jupyter Notebook

7 Python

4 SQL

5 R

5 Javascript

9 C++

11 C

8 MATLAB

5

10 Visual Studio Code (VSCode)

10 PyCharm

4 RStudio

10 Linear or Logistic Regression

10 Decision Trees or Random Forests

9 Gradient Boosting Machines

11 Convolutional Neural Networks

## Python

## MySQL

## Scikit-learn

## TensorFlow

## Keras

## Xgboost

## PyTorch

## LightGBM

## Caret

## Google Cloud Platform (GCP)

## Amazon Web Services (AWS)

## Microsoft Azure

## Oracle Cloud

## TensorFlow Extended (TFX)

## MLflow

## KServe

## TorchServe

## ONNX Runtime

## General purpose image video tools

## Image classification

## Image segmentation methods

## Object detection methods

## Contextualized embeddings

## Word embeddings vectors

## Transformer language models

## Encoder-decoder models

## Qlik Sense

## Google Data Studio

## Tableau

## Microsoft Power BI

## Google Cloud BigQuery

## Oracle Database

## MongoDB

## Scipy

## NumPy

## Pandas

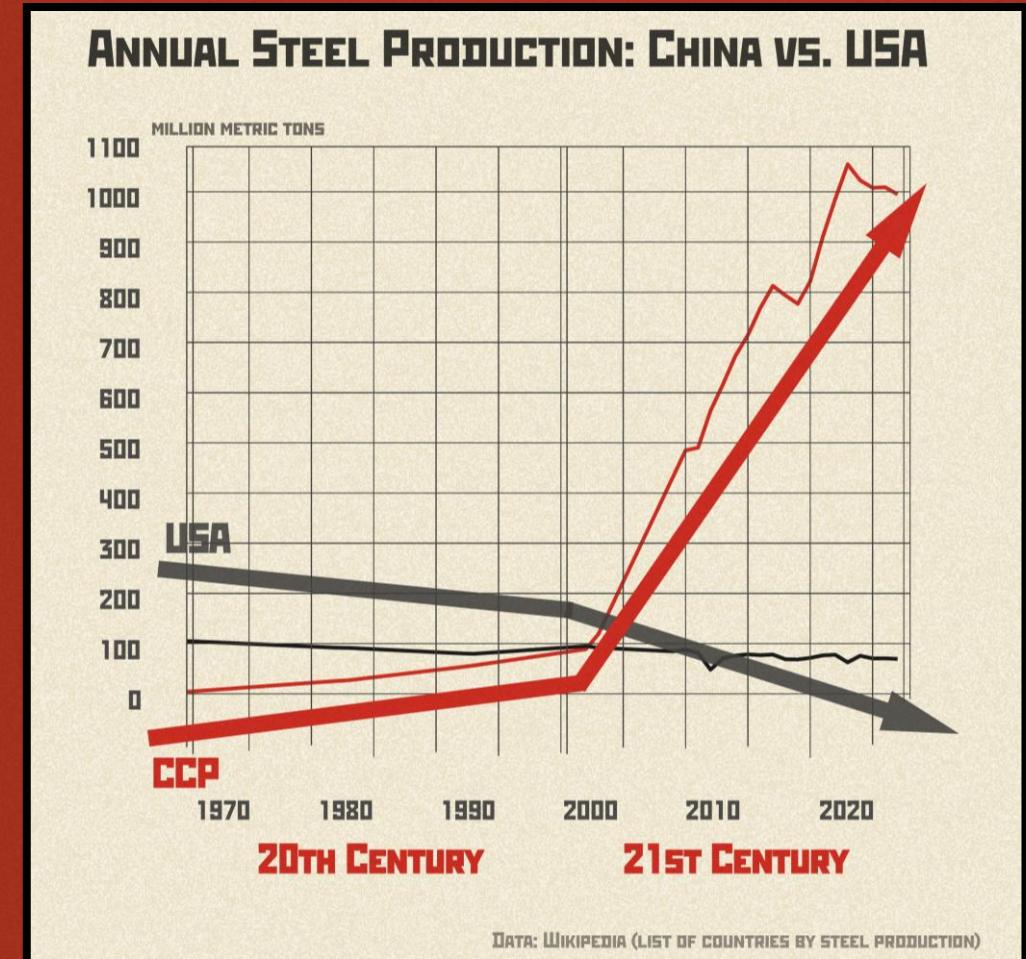
## PyTorch

## TensorFlow

# @DataCommissar



Visualization of Data in the Soviet Union





I want to  
go beyond using bar charts or repeating  
the same chart type

I want to  
go beyond just visualizing data like the  
GDP per capita data

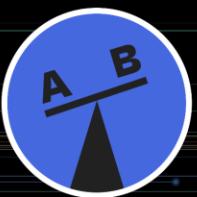
I want to  
go beyond just using the default themes  
and styles

# GETTING INSPIRED



# #30DayChartChallenge

April 2022 • 30 Days • 30 Charts • 5 Categories



## Comparisons

1. part-to-whole
2. pictogram
- 3. historical**
4. flora
5. slope
6. data day:  
OWID



## Distributions

7. physical
8. mountains
9. statistics
10. experimental
11. circular
12. theme day:  
The Economist



## Relationships

13. correlation
14. 3-dimensional
15. multivariate
16. environment
17. connections
18. data day:  
OECD



## Timeseries

19. global change
20. new tool
21. down/upwards
22. animation
23. tiles
24. theme day:  
Financial Times
25. trend
26. interactive
27. future
28. deviations
29. storytelling
30. data day:  
UN Population



## Uncertainties

Follow @30DayChartChall for more!

# Data: Dutch Loanwords



Showing 1 to 20 of 328 entries

Borrowed words	Relation	Donor languoid	Source word
<a href="#">Search</a>	--any--	<a href="#">Search</a>	<a href="#">Search</a>
<i>aanbidden</i>	immediate	Latin	<i>adorare</i>
<i>adelaar</i>	immediate	High German	<i>Adler</i>
<i>adres</i>	immediate	French	<i>adresse</i>
<i>alligator</i>	immediate	English	<i>alligator</i>
<i>altaar</i>	immediate	Latin	<i>altāre</i>
<i>anker</i>	immediate	Latin	<i>ancora</i>
<i>arts</i>	immediate	High German	<i>Artz</i>
<i>auto</i>	immediate	English	<i>auto(mobile)</i>
<i>baai</i>	immediate	French	<i>baie</i>
<i>baarmoeder</i>	immediate	Latin	<i>Latijn matrix</i>
<i>baby</i>	immediate	English	<i>baby</i>

[World Loanword Database \(WOLD\)](#)

# Data: Dutch Loanwords

Dutch Word	Source Language	Source Word	Meaning
 kaas	Latin	cāseus	cheese
 kaars	High German	kerza	candle
 rijk	Celtic	riks	rich
 haai	Old Norse	hár	shark

Searching for **text-heavy** chart  
inspirations

# Extinct Plants

## Africa



Ijeamaka Anyene, IUN declared extinct plants in Africa

# Loanwords in Dutch

Based on Nicoline van der Sijs' Loanwords in Dutch and the Dutch subdatabase of the World Loanword Database (WOLD). Of 1,513 core vocabulary items, 289 (19.1%) are **certain loans** and additional items are marked as **perhaps borrowed**. The main donor languages are French (36.7%), Latin (32.2%), German (14.2%), and English (7.6%). The database also records **calques words** (e.g., schoonmoeder 'mother-in-law').

^ shows words that have French Picard roots

+ shows words that have Low German roots (vs. High German)

@datacommissar

bambou / bambu  
bamboo / bongo  
eiland / eiland  
haar / hár  
kerk / kuri(a)kon  
dok / dok  
koefje / kahve  
nul / nulla  
rijf / ríjt  
rope / rópe  
thee / té  
zilver / arpu

banana / banana  
bataat / batata  
kakkerlak / cacalacca

pip / pipa  
pot / potus  
zijde / seda

Other

Portuguese

Romance

Celtic

coyote / coyote  
hangmat / hamaca  
kar / caña  
mais / maíz  
musket / mosquito  
pancho / pencho  
tabak / tabaco  
tapir / tapir  
toekan / tucán

trom / trum(m)e  
peddel / paddle  
phote / foto  
radio / radio

seks hebben / have sex

televisie / television

trouw / trouw

we / WC

yam / yam

Spanish

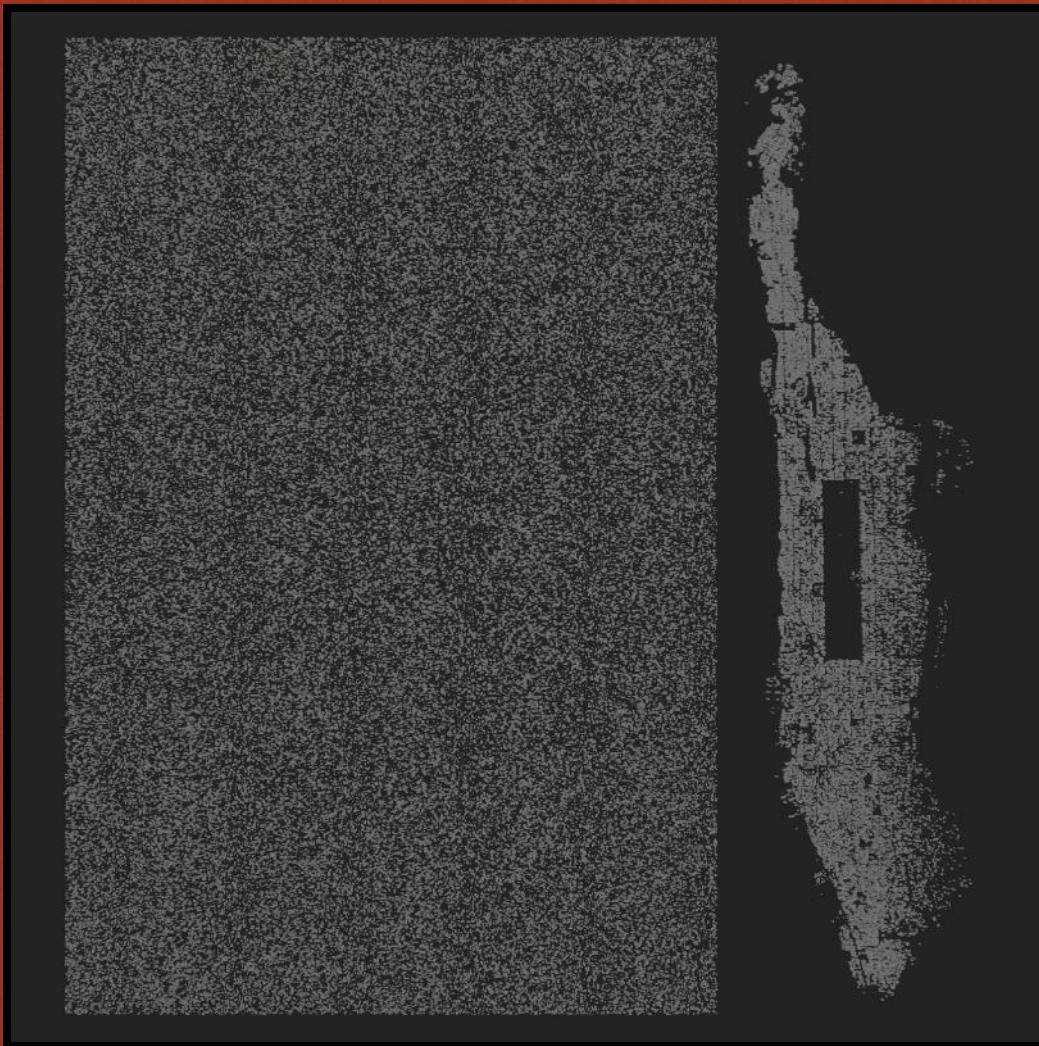
English

German

Latin

French

aanbidden / adorare	adres / adresse
altaar / altáre	buit / baie
anker / anker	batterij / batterie
baarmoeder / Latijn matrix	beest / bestie
bier / biber	bek / bec
bont / punctus	boeters / bocer^
boter / butter	bom / bombe
brie / brevis	bouffle / bouffle
bril / beryllus	cassave / cassave
cirkel / cercle	ceintuur / ceinture
demon / démon	direct / direct
dichter / dictator	dolfijn / dalfin, dauphin
dinsdag / dées Martis	dom / domine
doorn / doorn	elektricitet / electricity
donderdag / dies Iovis	familie / famille
duzendpool / milipedida	fee / fée
ezel / ézel	fife / flûte
granaat / granatum	fout / faute
kaas / cäesus	fruit / fruit
kameel / camélos	grootmoeder / grand-mère
kant / cantus	grootouders / grand-parents
kar / carrus	graat / grande
kat / catta	grat / gracie
keten / catena	juweel / joëel
keukken / cocina	jurk / journée
klam / clam	juwel / joël
loken / cocere	kadaver / cadavre
kop / cuppa	calebas / calebasse
koper / capo	kalim / calme
kort / curtus	kamp / camp^
krus / cruce	kariboe / caribou
lant / lantus	karotte / carotte
leeuw / león	kaos / colón
maandag / dies Lúnae	kleindochter / petite-fille
mantel / mantellum	kleinkind / petit-enfant
maret / maretus	kleinkind / petit-fils
mat / matras	kleur / couleur
medicijn / medicina	komijn / conin
metacretan / metacita	konijn / conin
mol / molaria	korai / corail
motel / mortarium	kramt / courant
motor / motor	krokodil / crocodile
nam / namus	kuiken / coquille
muur / murus	laagje / lagune
muur / murus	lampa / lanipe
noorderlicht / aurora borealis	machine / machine
oefen / offere	mannetje / manniere
olie / olum	manioek / manioc
oli / oliva	matras / materas
oogst / oestus	mees / meesje
paal / pálus	minister / ministre
paar / par	moeras / marasch
paard / paravredus	moskee / mosque
paard / parvatura	mozaïek / mozaïque
pon / panha	ocean / ocean
pell / pellis	olifant / olifant
pen / pena	papier / paperna
penis / pénis	papgeai / paggaï
peper / piper	papier / papier
persoon / persona	pissen / pissier
pil / pilus	pat / patre
pijn / pena, poena	plank / plankeΛ
pil / pilula	plat / plat
pin / pinna	poldertje / police
planten / plantare	pompon / pompon
pols / pulsera	post (2) / poste
poort / porta	president / président
post (1) / postis	pro / pro
preken / praedicare	proeven / prover
priester / prestre	prostituee / prostituée
prosternen / prostrate	resten / reste
pus / pùs	rij / rij
regeren / regere	rivier / rivière
reiger / reiger	rollen / rouler
reiger / reiger	rot / rond
ruggengraat / spina dorsi	roos / rose
hinderlaag / Hinterthal	ros / roch^
hutte / hutte	savanne / savanne
kaars / kerza	schotella / schotella
kachel / Kachelofen	scrijven / scribera
kanarie / kanarie	sláaf / slavus
kapot / kaputti	slenter / slenter
langzaam / langsam	sorghum / sorghum
mededijken / Mitleid	specie / species
reiger / reiger	speld / spinula
oorzaak / Ursache	spontane / spontane
pech / Pech	strata / strata
reinheit / reinher	studeren / studere
reis / reis	tempel / tempel
siegel / Zierat	testikels / testiculus
sintel / sintel	toren / turrem
sluier / Schleier	troon / trone
spindele / spindel	vaderland / patria
spits / Spitzer	vagina / vagina
stotteren / Stottern	venster / fenestra
strof / strofe	voet / voet
trom / trum(m)e	vrijdag / dies Veneris
trommel / Trommel	vulva / vulva
phote / foto	woensdag / dies Mercurii
radio / radio	zak / saccus
seks hebben / have sex	zaterdag / dies Saturni
televisie / television	zegel / zegare
trouw / trouw	wikkelen / wickelen
we / WC	zeker / securare
yam / yam	zondag / diés sólis
zich herinneren / sich erinnern	zich haasten / se haster



Reuters: Gaza in rubble and ruin

## Overlaying Destruction: Visualizing Gaza's Damage in Terms of Amsterdam

As of September 2024, UNOSAT reported over 128,000 buildings in Gaza destroyed or damaged due to Israeli attacks: 52,564 destroyed, 18,913 severely damaged, and 56,710 moderately damaged. This plot illustrates the scale of destruction Amsterdam was attacked the same manner as what it has done to Gaza. Such a scenario would leave nearly all buildings damaged or completely destroyed

■ Destroyed ■ Severely damaged ■ Moderately damaged



# #TidyTuesday Contributions

A screenshot of a GitHub repository page for "z3tt/TidyTuesday". The repository is public and has 1 branch and 0 tags. It contains 606 commits from z3tt, last updated 2 years ago. The repository has 144 forks. The README file contains a section titled "#TidyTuesday Contributions" with social media sharing icons and a "Buy me a coffee" button.

z3tt/TidyTuesday Public

<> Code Issues 1 Pull requests Actions Security Insights

main 1 Branch 0 Tags Go to file <> Code About

z3tt Update LinkedIn Handle 0dceb27 · 2 years ago 606 Commits

R rm DS\_Store 2 years ago

README.md Update LinkedIn Handle 2 years ago

TidyTuesday.Rproj 2019/17 6 years ago

README #TidyTuesday Contributions

Buy me a coffee

144 forks Report repository

No releases published

No packages published

z3tt Cédric Scherer

gkaramanis Georgios Karamanis

charts + code

# Datawrapper Data Vis Dispatch

z3tt / TidyTuesday Public

Datawrapper | Blog Login Start creating

September 2nd, 2025 by Rose Mintzer-Sweeney 10 min

 Data Vis Dispatch, August 26: Africa map, nations' energy, and more

August 26th, 2025 by Jonathan Dahle 1 min

Tuesday visualization created in R  
rstats  
y-challenge

# ~6k visualizations

 Data Vis Dispatch, August 12: Burning Europe, interactive maps, and tariff analyses

August 12th, 2025 by Jonathan Dahle 1 min

 Data Vis Dispatch, August 5: Germany, Texas, and Coffee

August 5th, 2025 by Jonathan Dahle 7 min

# Flowingdata.com

z3tt / TidyTuesday Public

Datawrapper | Blog Login Start creating

**FLOWINGDATA** Membership Books Projects Learning Newsletters Member Login

Data Underload / occupation, salary, work

## Salary and Occupation

Support an independent FlowingData, and get extra visualization goodness.

BECOME A MEMBER →

# ~9k visualizations



The median salary for full-time workers in the United States was \$49,500, based on estimates from the Bureau of Labor Statistics in 2024. However, salaries vary by occupation. These charts show the spread.

August 21, 2025  
Visualization editing

August 14, 2025  
Careless chart mistakes

August 7, 2025  
Familiar chart advantages

*Second Edition*

 Visualize This: The FlowingData Guide to Design, Visualization, and Statistics (2nd Edition)  
New tools, refined process.  
Order: [Amazon](#) / [Bookshop](#)

# Pinterest

z3tt / TidyTuesday Public

Datawrapper | Blog | Login | Start creating | Member Login

FLOWINGDATA Membership Books Projects Learning Newsletters

All Data visualization portrait ideas helios

Search

Can't even count them

Tuesday  
ilization  
ated in R

rstats  
y-challenge

# LinkedIn

The image is a collage of several screenshots from different websites, all related to data visualization and tidy data. At the top, there's a screenshot of a blog post titled 'z3tt / TidyTuesday' on Datawrapper, featuring a header with 'Datawrapper' and 'Blog' and a navigation bar with 'Login' and 'Start creating'. Below it is a screenshot of the FlowingData website, which has a dark blue header with the 'F' logo and the word 'FLOWINGDATA' in white. The main content area shows a search bar and various navigation links like 'Home', 'My Network', 'Jobs', 'Messaging', 'Notifications', 'Me', 'For Business', and 'Learning'. To the right of the FlowingData screenshot is a vertical column of text snippets from a Tuesday Tidy Visualization challenge in R, including 'Tuesday', 'Visualization', 'dated in R', 'rstats', and 'challenge'. The bottom half of the collage is dominated by a large screenshot of the LinkedIn interface. On the left, there's a sidebar for 'My items' showing 'My jobs' (268) and 'My learning' (69). The main feed shows a post from 'Veerle Eeftink - van Leemput' (2nd connection) about R and shiny. Below it are two more posts: one from 'Cara Thompson' (1st connection) about accessibility at the RSS conference, and another from 'Yan Holtz' (2nd connection) about staying up-to-date on LinkedIn.

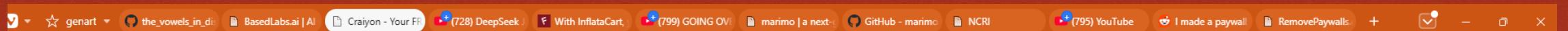
## Didn't even try counting

# Bluesky

The image shows a collage of several social media and data visualization platforms, illustrating the concept of Bluesky as a competitor or alternative to these services.

- LinkedIn:** A screenshot of the LinkedIn homepage featuring a search bar and navigation links for Home, My Network, Jobs, Messaging, Notifications, Me, For Business, and Learning.
- FlowingData:** A screenshot of the FlowingData website, which is a blog about data visualization. It features a header with "z3tt / TidyTuesday" and "Public" status, and a main content area with a post about "Tuesday visualization generated in R".
- Pinterest:** A screenshot of the Pinterest homepage with a search bar and navigation links for Home, My Network, Jobs, Messaging, Notifications, Me, For Business, and Learning.
- Datawrapper:** A screenshot of the Datawrapper website, showing a dashboard with various data visualization tools and resources.
- Bluesky:** The central focus is a screenshot of the Bluesky mobile application. It displays a feed of posts, one of which is a "Trending" post featuring a "The Wizard of Oz" theme with images of Dorothy, Toto, the Scarecrow, and the Cowardly Lion. The Bluesky interface includes a sidebar with navigation links like Home, Explore, Notifications, Chat, and Feeds, and a "GETTING STARTED" guide with tips like "Like 10 posts" and "Follow 7 accounts".

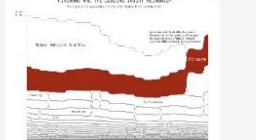
**See above**



# Solution 3: NOTION

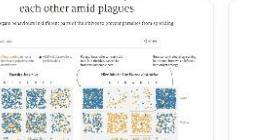
Main Page / Data Journalism / Interesting Charts

Edited Jan 10 Share ★

  
Deaths by Firearm, Compared Against Injury-Related Deaths

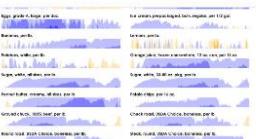
  
2000 1488 5.544t  
albums vs streaming

  
the city forest

  
each other amid plagues

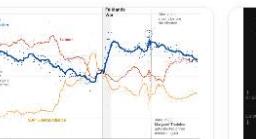
  
Just like modern humans, honeybees avoid each other amid plagues

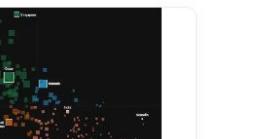
  
New page

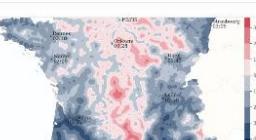
  
Horizon Graphs

  
Visualizing the EU's Energy Dependency

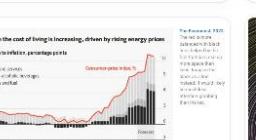
  
A table set for fasting

  
Falklands effect

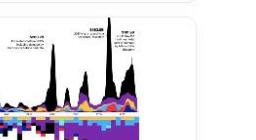
  
In the climate crisis, vulnerable countries bear the least responsibility

  
New page

  
New page

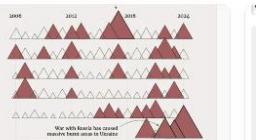
  
New page

  
Tree rings to compare life expectancy in your state

  
Billion dollar natural disasters

  
New page

  
New page

  
New page

  
FATHER WOMAN MOTHER  
SINCE POWER WHITE DEATH  
PASSED LEAVE NATURE CHILD WOMEN  
REPLIED EARTH THOUSAND SPIRIT TRUTH  
HUMAN RETURNED AROUND SENSE FOLLOWED

  
New page





Newsletters,  
twitter, LinkedIn ...

Me trying to keep  
up with new  
visualizations

1,280



*"I want AI to do my laundry and dishes so that I can do art and writing, not for AI to do my art and writing so that I can do my laundry and dishes."*

Author and videogame enthusiast  
(although both)



**85%** Time spent on actually creating charts

**10%** Time spent searching in the saved charts collection

**5%** Time spent saving charts to Notion

What if, instead of scrolling, I  
built a system that can  
**semantically** retrieve charts?

# Examples

I'm looking for charts **house ownership rates**

I'm looking for charts that show **proportions or distributions**

I'm looking for **circular bar plots**

I'm looking for charts that have **diverging legends**



problems in my life



**IT'S  
DEMO  
TIME**

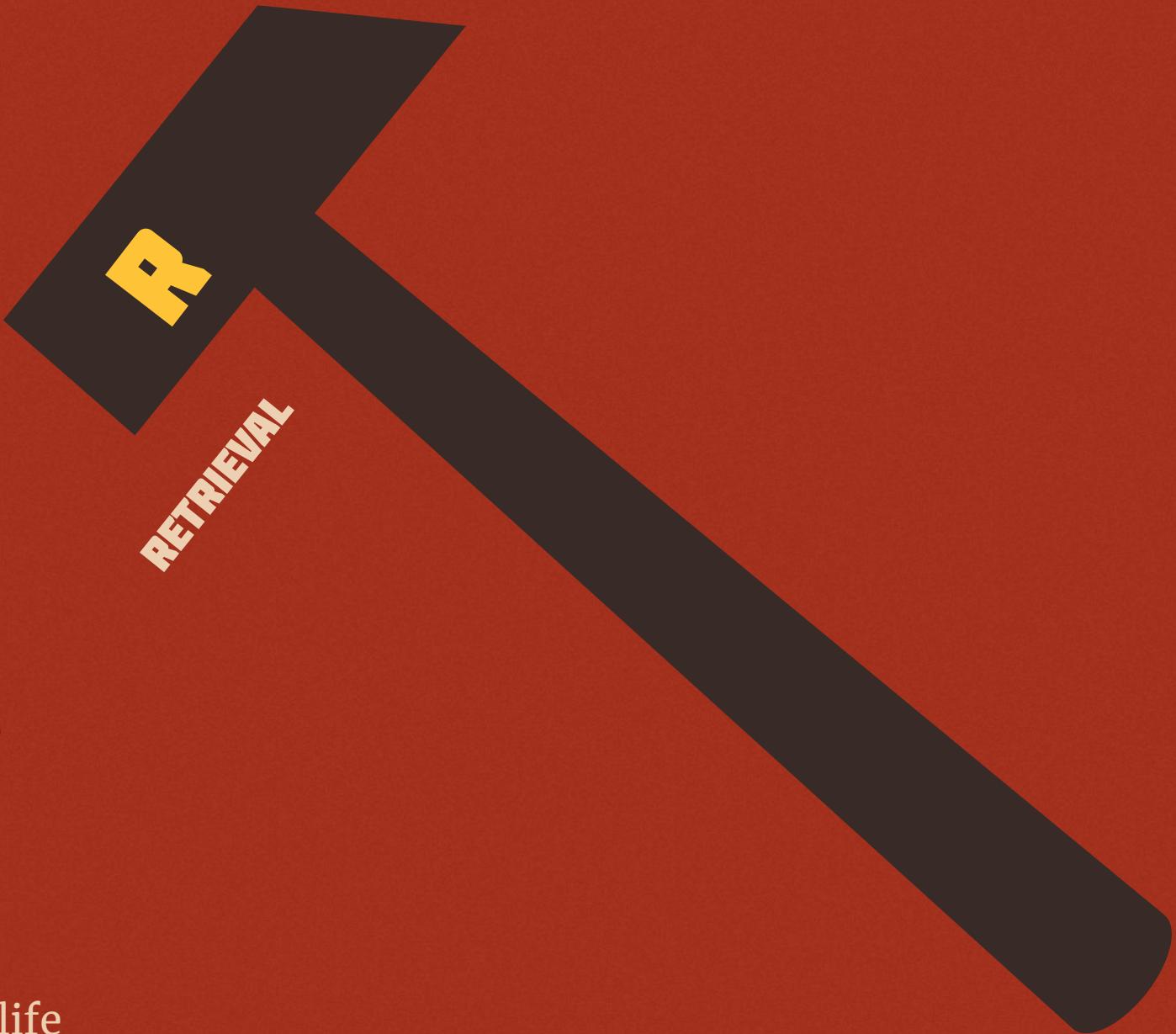
**GO TO  
[MYNEXTCHART.COM](http://MYNEXTCHART.COM)**



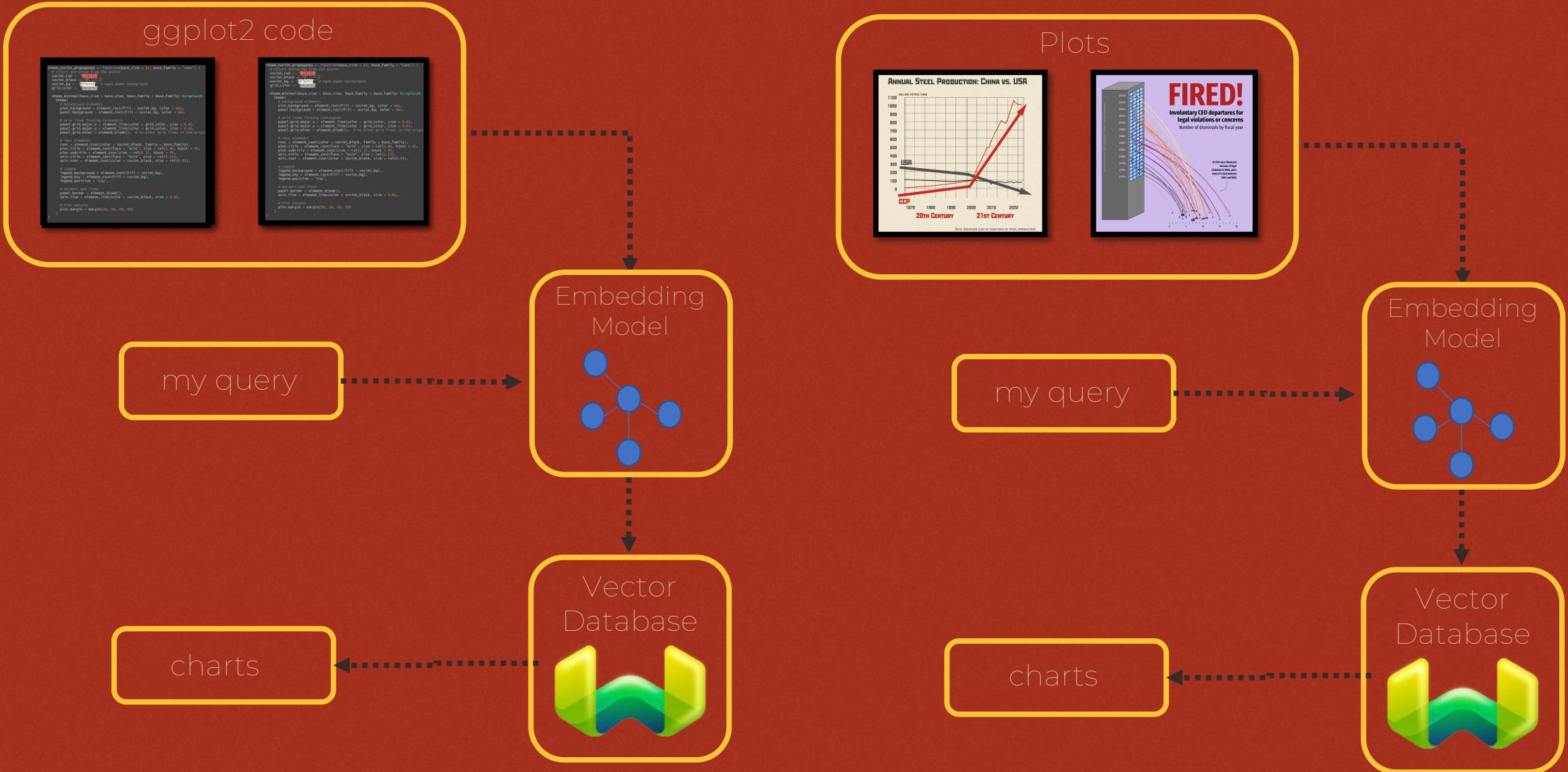
# **TECHNICAL IMPLEMENTATION**



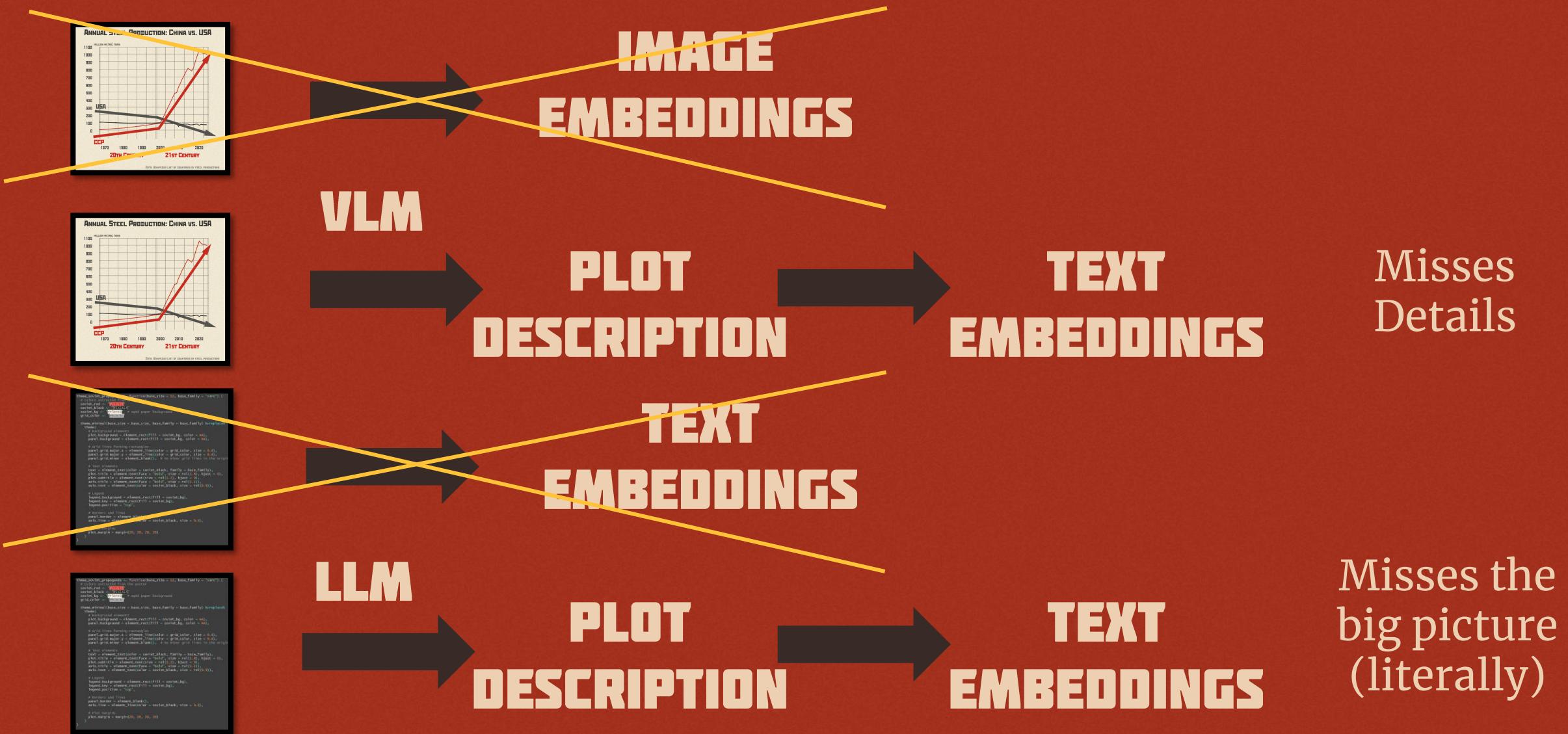
problems in my life



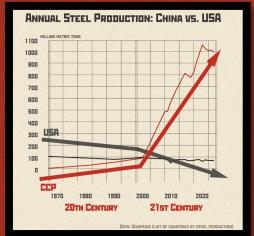
# Retrieving Charts



# From Graphs to Vectors



# From Graphs to Vectors



```
ggplot(data=ggplot2::gapminder::gapminder, aes(x=year, y=steel)) +  
  geom_line(aes(color=country), size=1, lineend="round") +  
  geom_point(aes(color=country), size=1, lineend="round") +  
  scale_color_manual(values=c(CHINA="red", USA="black")) +  
  theme_minimal() +  
  theme(panel.grid.major.x=element_rect(fill="white", color="black", size=0.5),  
        panel.grid.minor.x=element_rect(fill="white", color="black", size=0.5),  
        panel.grid.major.y=element_rect(fill="white", color="black", size=0.5),  
        panel.grid.minor.y=element_rect(fill="white", color="black", size=0.5),  
        plot.background=element_rect(fill="white", color="black", size=0.5),  
        plot.title=element_rect(fill="white", color="black", size=0.5),  
        plot.subtitle=element_rect(fill="white", color="black", size=0.5),  
        plot.caption=element_rect(fill="white", color="black", size=0.5),  
        axis.ticks=element_rect(fill="white", color="black", size=0.5),  
        axis.title=element_rect(fill="white", color="black", size=0.5),  
        axis.subtitle=element_rect(fill="white", color="black", size=0.5),  
        axis.text=element_rect(fill="white", color="black", size=0.5),  
        legend.title=element_rect(fill="white", color="black", size=0.5),  
        legend.text=element_rect(fill="white", color="black", size=0.5),  
        panel.border=element_rect(fill="white", color="black", size=0.5),  
        panel.spacing=margin(20, 20, 20, 20))
```

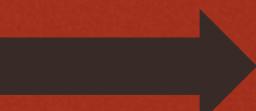
Only plots made  
by ggplot2 from  
Github



VLM



QWEN-VL-MAX



PLOT  
DESCRIPTION



TEXT

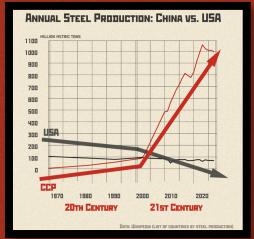
EMBEDDINGS



MISTRAL  
AI

Mistral Embeddings

# From Graphs to Vectors



```
ggplot(data=ggplot2::gapminder::gapminder, aes(x=year, y=pop, color=country)) +  
  geom_point() +  
  geom_line()  
  # Add a legend  
  scale_color_manual(values=c("USA" = "#E69138", "China" = "#C8512E"))  
  # Set the background color to white  
  theme_void() +  
  theme(plot.background = element_rect(fill = "white", color = "white", size = 0),  
        panel.background = element_rect(fill = "white", color = "white", size = 0),  
        grid = element_rect(color = "white", fill = "white"),  
        plot.title = element_rect(size = 10, color = "black", fill = "white", fontweight = "bold"),  
        plot.subtitle = element_rect(size = 10, color = "black", fill = "white", fontweight = "bold"),  
        plot.caption = element_rect(size = 10, color = "black", fill = "white", fontweight = "bold"),  
        axis.title = element_rect(size = 10, color = "black", fill = "white", fontweight = "bold"),  
        axis.subtitle = element_rect(size = 10, color = "black", fill = "white", fontweight = "bold"),  
        axis.ticks = element_rect(size = 10, color = "black", fill = "white", fontweight = "bold"),  
        axis.line = element_rect(size = 1, color = "black", fill = "white", fontweight = "bold"),  
        plot.margin = margin(20, 20, 20, 20))
```

Only plots made by ggplot2 from Github



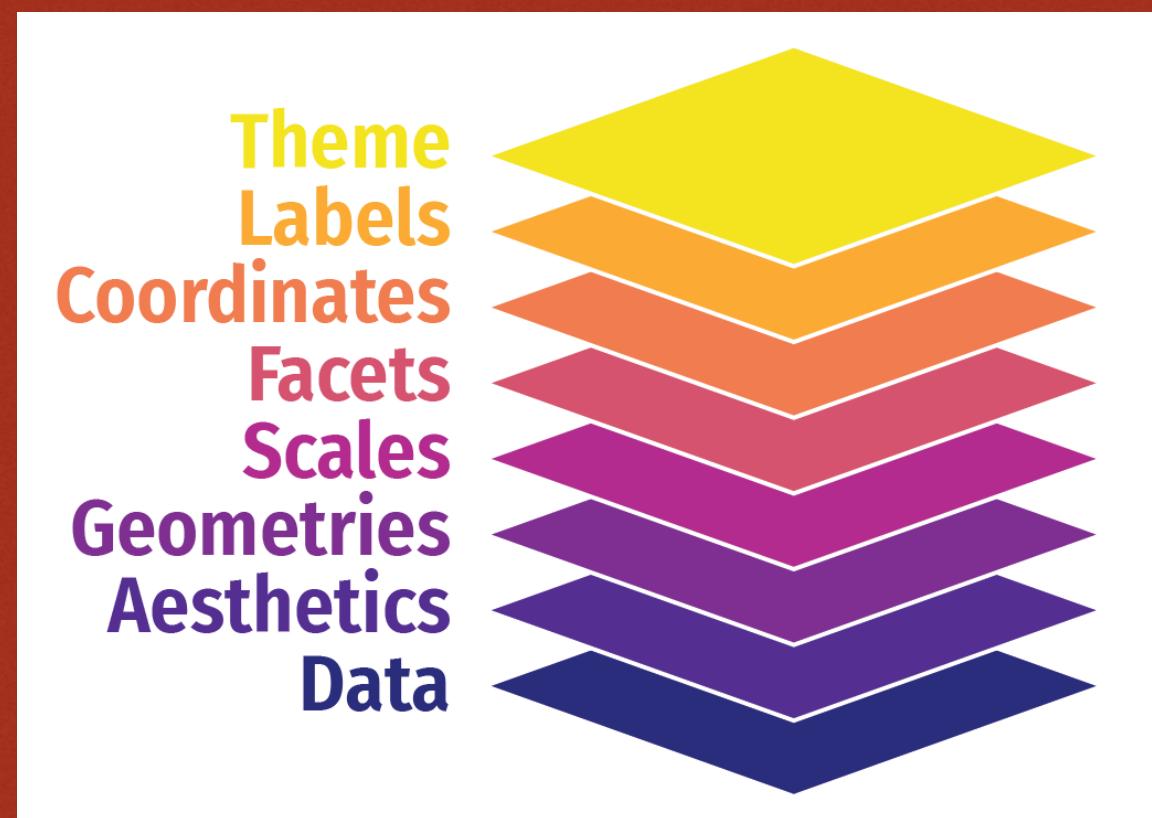
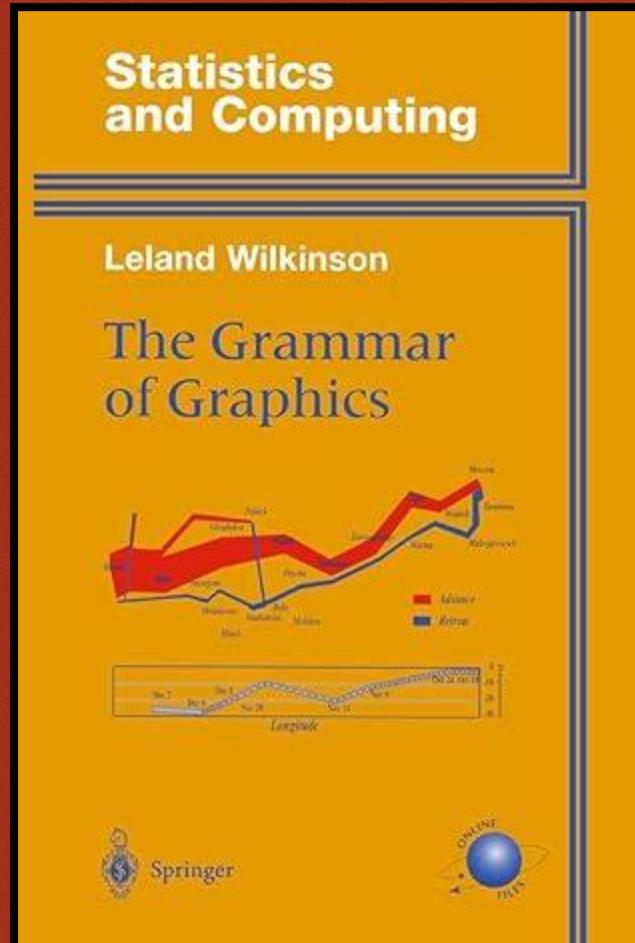
PLOT  
DESCRIPTION

TEXT  
EMBEDDINGS



Mistral Embeddings

# Grammar of Graphics



Data Visualization by Andrew Heiss

# Grammar of Graphics

Component	ggplot2 (R)	plotnine (Python)	Explanation
Data	ggplot(data)	ggplot(data)	The raw data that you want to visualize (initializing a plot).
Aesthetics	aes()	aes()	The mapping between variables and visual properties.
Geometries	geom_*	geom_*	The geometric shape of a layer representing the data.
Statistics	stat_*	stat_*	The statistical transformation of a layer applied to the data.
Scales	scale_*	scale_*	The representation of mapped aesthetic attributes.
Coordinate System	coord_*	coord_*	The transformation to map data coordinates into the plot plane.
Facets	facet_*	facet_*	The arrangement of the data into a set of small multiples.
* Themes	theme()	theme()	<i>The overall visual defaults of non-data elements of the graphic.</i>

Lets' look at  
the theme

# IMPACT OF AI LANGUAGE MODELS ON VARIOUS JOB CATEGORIES

DISTRIBUTION OF LANGUAGE MODELING EXPOSURE ACROSS JOB CATEGORIES, SHOWING WHICH OCCUPATIONS RELY MOST HEAVILY ON COMMUNICATION AND LANGUAGE ABILITIES THAT LLMs CAN POTENTIALLY AUTOMATE OR AUGMENT, WITH EXAMPLES OF THE MOST AND LEAST IMPACTED ROLES WITHIN EACH JOB CATEGORY

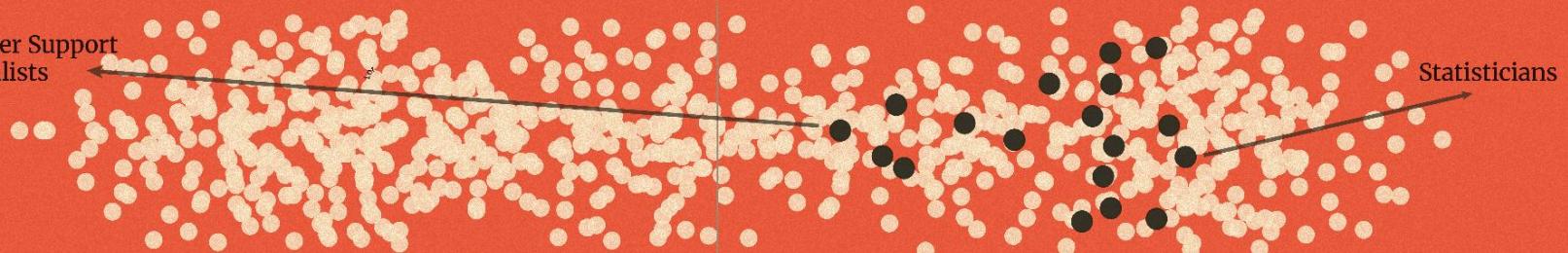
Less Impact

More Impact

## COMPUTER AND MATHEMATICAL

Computer User Support Specialists

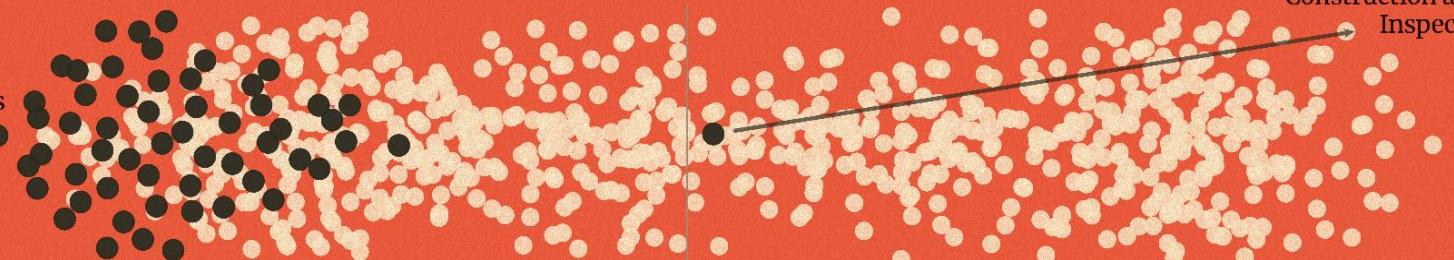
Statisticians



## CONSTRUCTION AND EXTRACTION

Reinforcing Iron and Rebar Workers

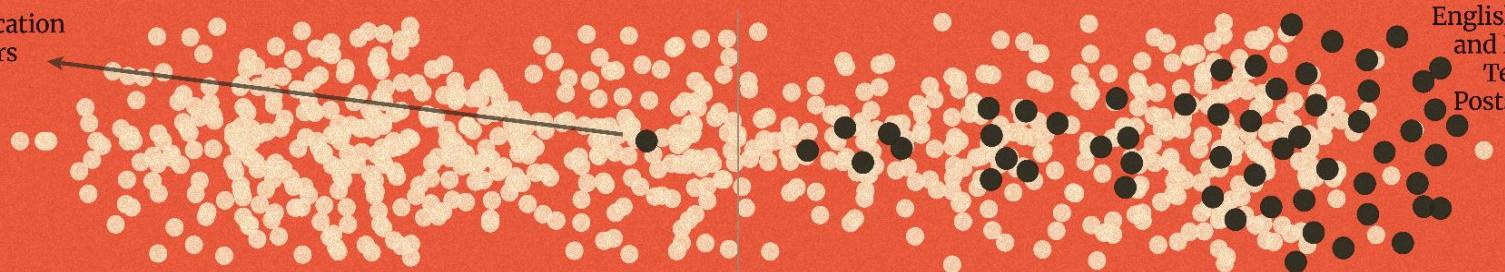
Construction and Building Inspectors



## EDUCATION, TRAINING, AND LIBRARY

Special Education Teachers

English Language and Literature Teachers, Postsecondary



# Theme

Background color: #E95A3C

Title: Center aligned

Subtitle: Center aligned

Title font: **MOLOT**

Subtitle font: **MOLOT**

Font color (black),

Font style (**bold**)

Legend: None

Grids: None

Axis: None

etc.

## IMPACT OF AI LANGUAGE MODELS ON VARIOUS JOB CATEGORIES

DISTRIBUTION OF LANGUAGE MODELING EXPOSURE ACROSS JOB CATEGORIES, SHOWING WHICH OCCUPATIONS RELY MOST HEAVILY ON COMMUNICATION AND LANGUAGE ABILITIES THAT LLMS CAN POTENTIALLY AUTOMATE OR AUGMENT, WITH EXAMPLES OF THE MOST AND LEAST IMPACTED ROLES WITHIN EACH JOB CATEGORY

Less Impact

More Impact

### COMPUTER AND MATHEMATICAL

Computer User Support Specialists

Statisticians

### CONSTRUCTION AND EXTRACTION

Reinforcing Iron and Rebar Workers

Construction and Building Inspectors

### EDUCATION, TRAINING, AND LIBRARY

Special Education Teachers

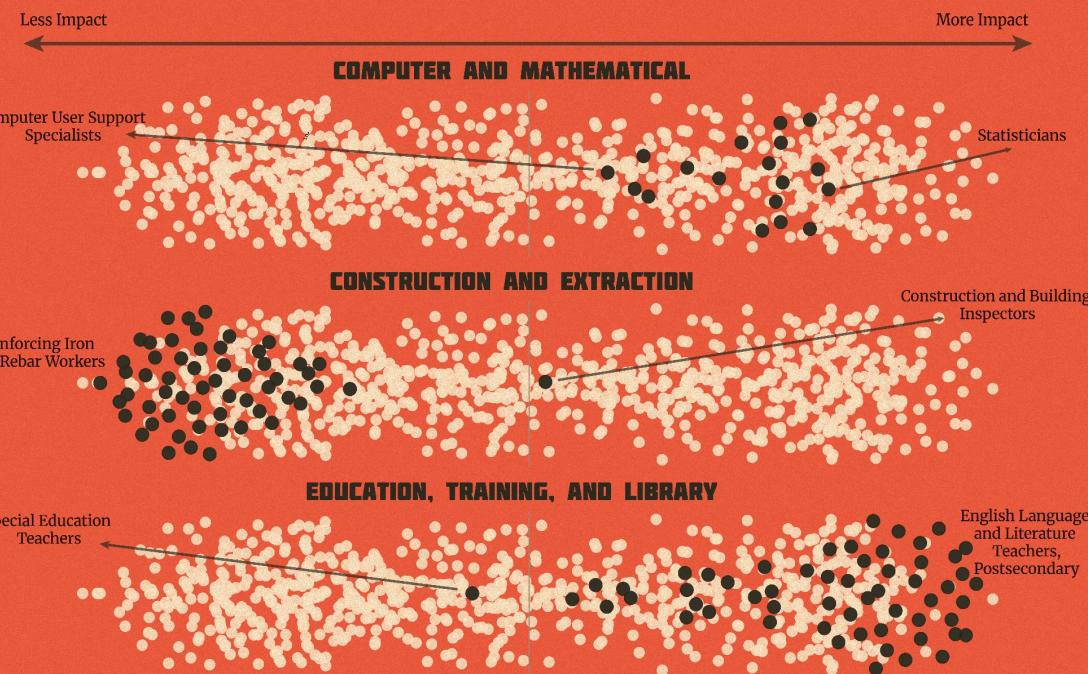
English Language and Literature Teachers, Postsecondary

# prompt\_v1.txt

```
1 You're an expert in analyzing ggplot2 visualizations and R code. Your goal is to provide a comprehensive technical breakdown of the following R code that uses ggplot2 to create a visualization.
2 Format your response with these numbered sections separated by %%%:
3 0. **Plot Category & Purpose**
4     - Understand and describe what this plot represents in high-level terms, for example: "This code creates a visualization showing the distance between each US state's capital city and its largest city.
5     - Describe the type of data being visualized and the overall purpose of the plot.
6     - Describe the visualization approach in detail, including the type of plot, the data being used, and any specific features or patterns that stand out.
7     - Target audience analysis (who would benefit most from this visualization)
8     - Visualization purpose (exploratory vs explanatory)
9     - Alternative visualization approaches that could present the same data
10    - The logic of the data visualization
11    - Based on the categories of the previous editions of #30DayChartChallenge, identify which categories apply (separated by commas): Experimental, Pop Culture, Stats & Models, Comparison, Distribution,
12
13 1. **Data Source & Preparation**
14     - Data source identification and retrieval methods
15     - Key data transformations before plotting
16     - Data cleaning and filtering techniques applied
17     - Variable creation and manipulation
18     - External data merges or joins
19
20 2. **Title & Text Elements**
21     - Plot title, subtitle, and caption with exact text
22     - Annotations and direct labels (geom_text/label, annotate()) with positioning parameters
23     - Text formatting specifications (size, color, face)
24     - Contextual information provided through text
25
26 3. **Geometry & Mappings**
27     - All geometry types used (list each geom with full parameters)
28     - Variable mappings for each geometry (what maps to x, y, color, fill, size, etc.)
29     - Layer order and drawing precedence
30     - Aesthetic mapping inheritance patterns
31
32 4. **Statistical Transformations & Coordinates**
33     - Faceting implementation (facet_wrap/grid variables and scales)
34     - Coordinate system details (coord_flip, coord_polar, limits, expansions)
35     - Statistical transformations (stat_summary, binning, smoothing parameters)
36     - Scale transformations (log, sqrt, etc.)
37
38 5. **Theme Architecture**
39     - Base theme template (theme_minimal, theme_bw, etc.)
40     - Complete theme() chain with exact parameter values
41     - Notable theme overrides categorized by element type (text, line, rect)
42     - Custom margin and spacing adjustments
43     - Overall aesthetic intention of the theme choices
44
45 6. **Color Specifications**
46     - Named color palette or custom scale functions with hex codes
47     - Named background color and fill with hex codes
48     - Manual color/fill values with hex codes
49     - Transparency settings (alpha values)
```

## IMPACT OF AI LANGUAGE MODELS ON VARIOUS JOB CATEGORIES

DISTRIBUTION OF LANGUAGE MODELING EXPOSURE ACROSS JOB CATEGORIES, SHOWING WHICH OCCUPATIONS RELY MOST HEAVILY ON COMMUNICATION AND LANGUAGE ABILITIES THAT LLMs CAN POTENTIALLY AUTOMATE OR AUGMENT, WITH EXAMPLES OF THE MOST AND LEAST IMPACTED ROLES WITHIN EACH JOB CATEGORY



%%%  
0. Plot Category & Purpose

- **High-level description:**

This visualization shows how **AI language models (LLMs)** are expected to impact different job categories. Each dot represents an occupation within that category, positioned along a horizontal axis from "Less Impact" to "More Impact." Specific occupations are highlighted and annotated to illustrate extremes (least and most affected).

- **Type of data:**

Occupational data, impact score (continuous), categorical job groups, and highlight markers.

- **Visualization approach:**

A faceted strip plot / dot scatter plot:

- X-axis = degree of LLM impact.
- Y-axis = different job categories arranged in stacked panels.
- Cream dots = all occupations.
- Black dots = highlighted specific jobs (annotated with arrows and labels).

- **Geometry & Mappings**

- **Geoms:**

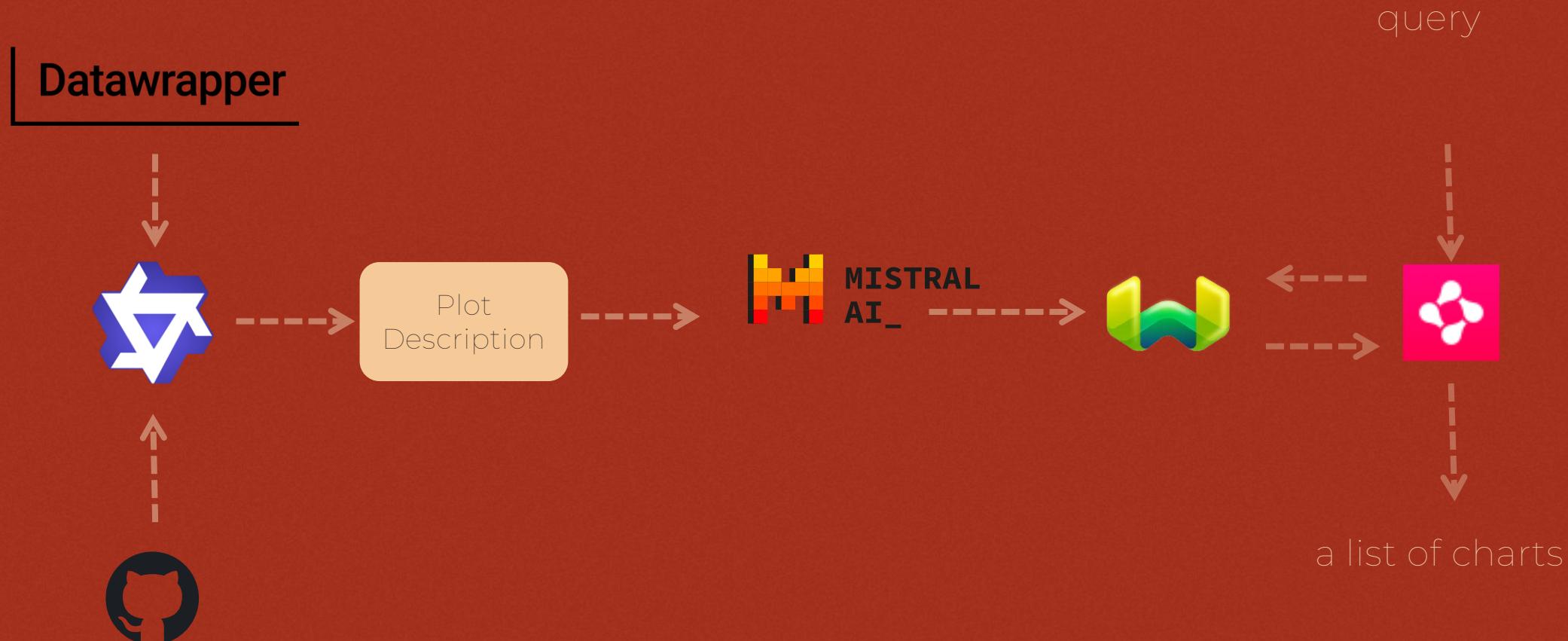
- `geom_point(alpha=0.8, size=2)` → scatter of occupations.
- `geom_point(size=2.5, color="black")` → highlight jobs.
- `geom_text(hjust/vjust adjustments)` → occupation annotations.
- `geom_segment(arrows=arrow(), linetype="solid")` → connecting arrows.

- **Mappings:**

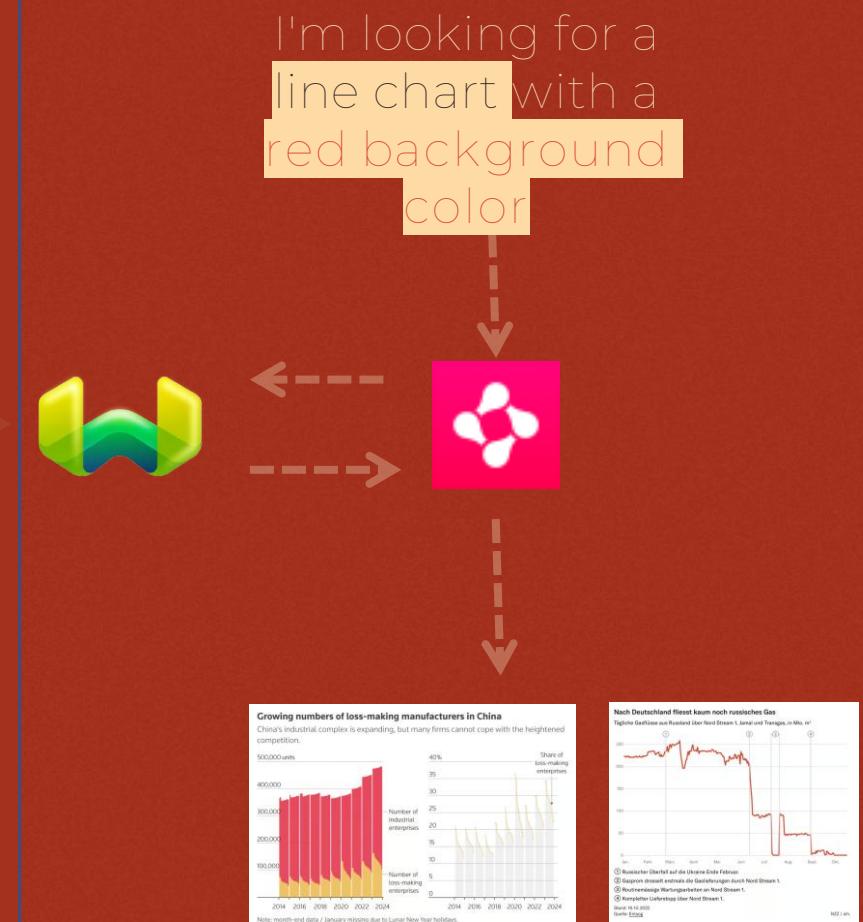
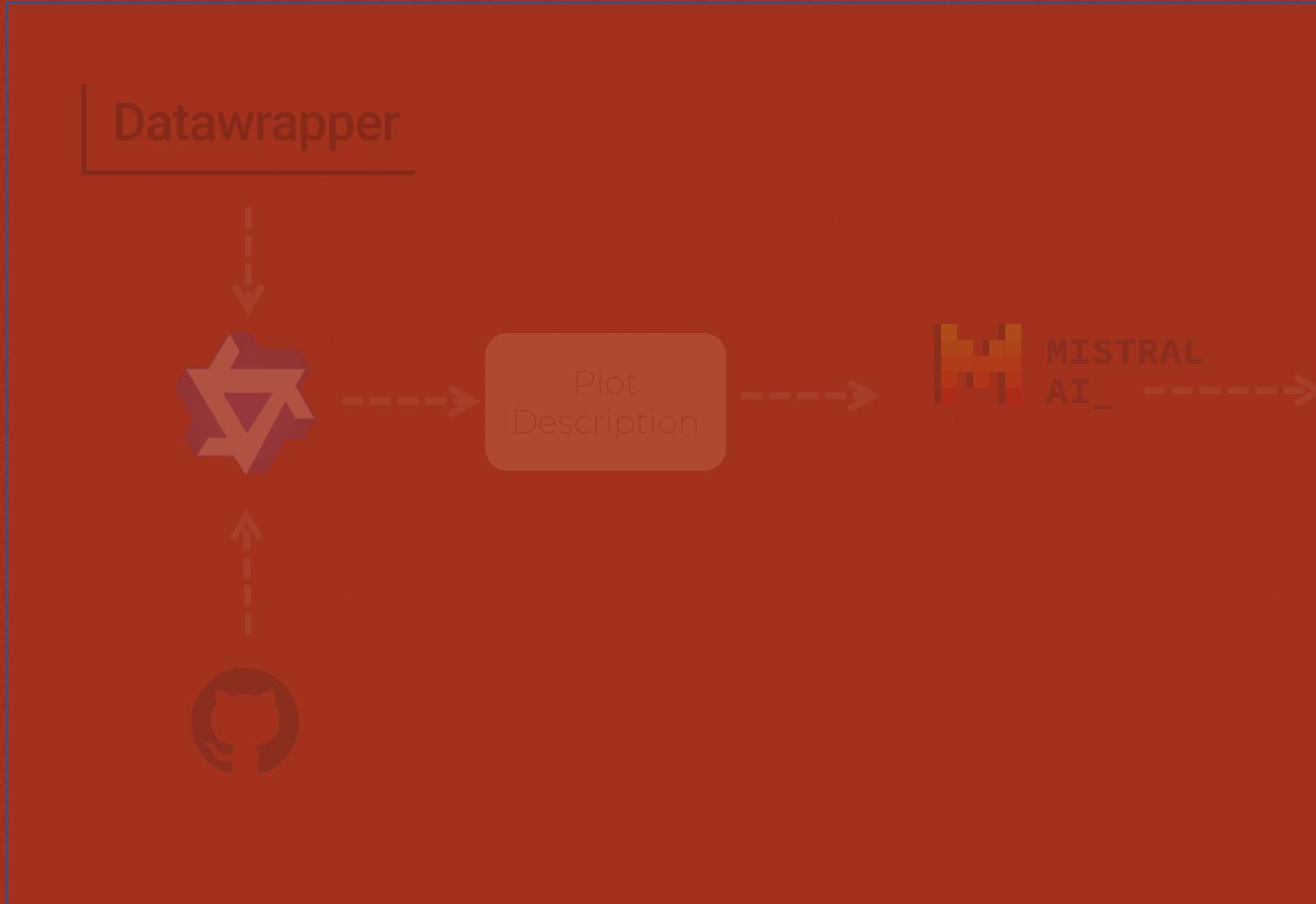
# **MODE 1**

## **VANILLA/HYBRID RETRIEVAL**

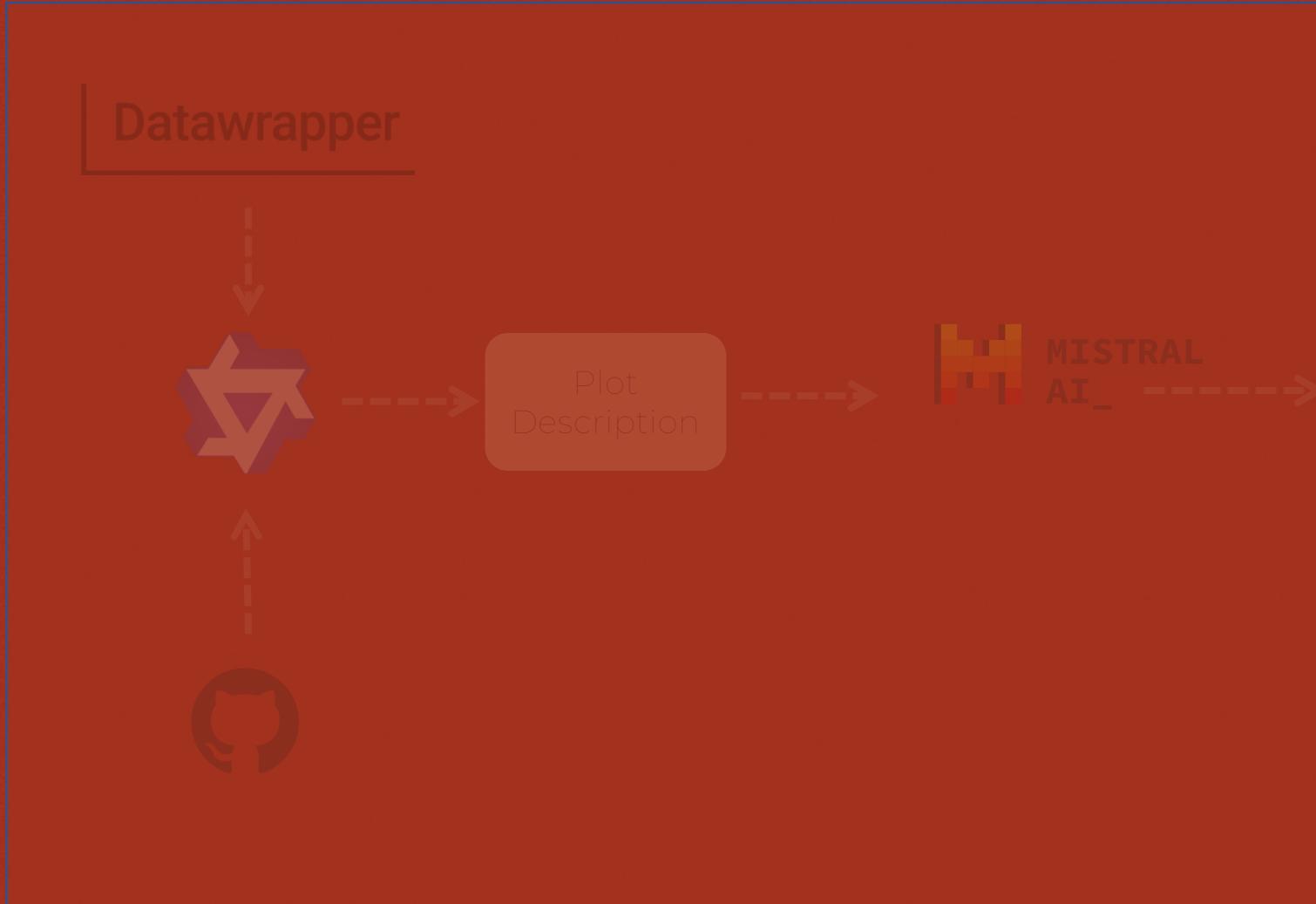
# THE PROBLEM WITH VANILLA RETRIEVAL



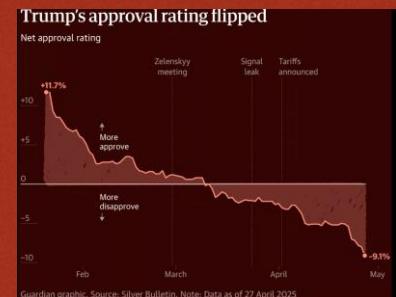
# THE PROBLEM WITH VANILLA RETRIEVAL



# THE PROBLEM WITH VANILLA RETRIEVAL



I'm looking for a  
line chart with a  
red background  
color



# The “Semantic Soup” Problem



# **MODE 2**

## **DECONSTRUCTING VISUALIZATIONS WITH TOOL CALLING**



Plot type =  
line chart

BG  
color = red

Deconstructor

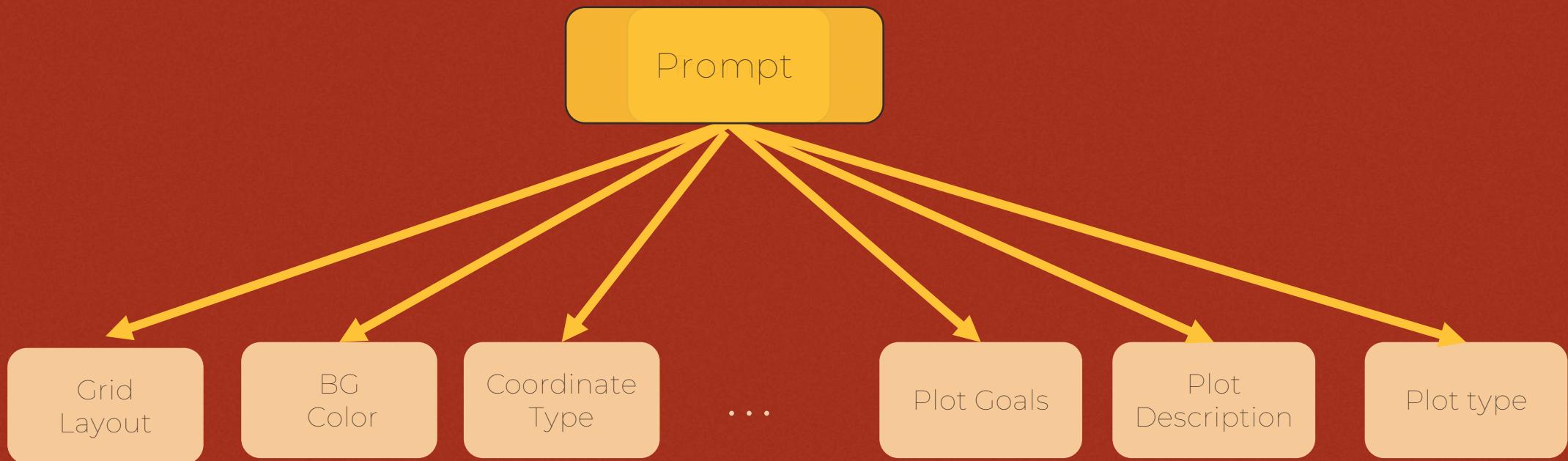


I'm looking for a  
line chart with a  
red background  
color



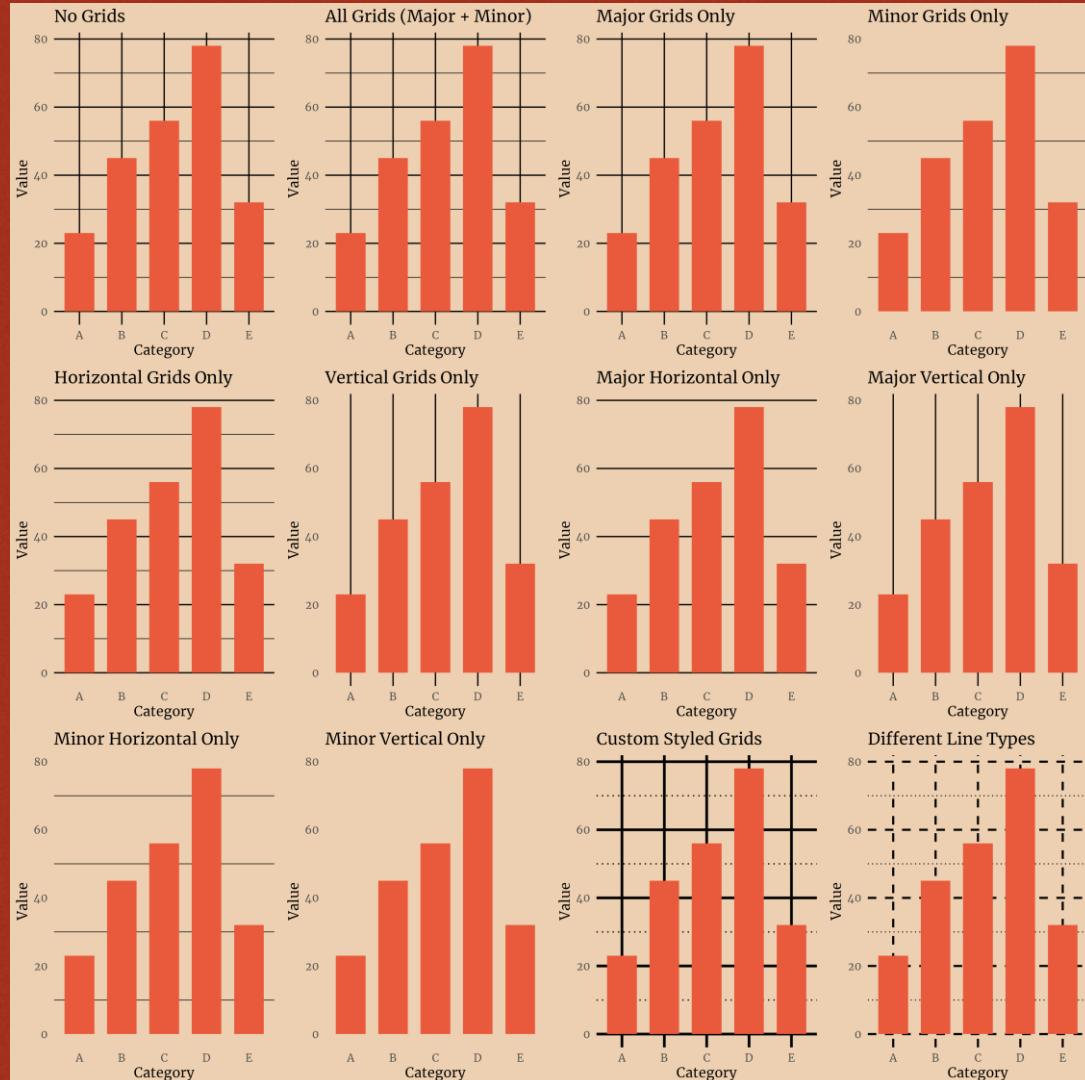
a list of charts

# Deconstructing the Prompt

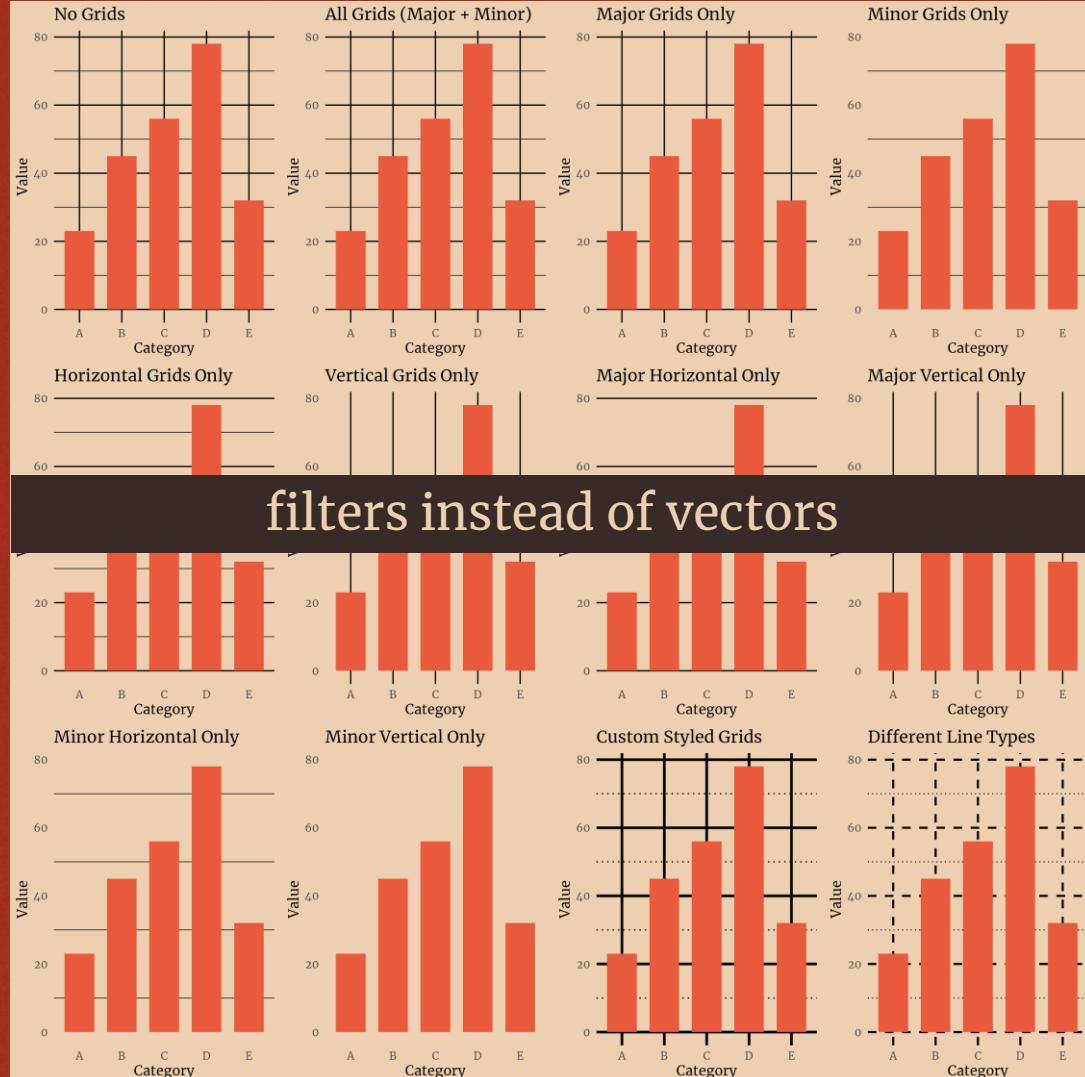


Each element stored as a vectorized field in the vector DB

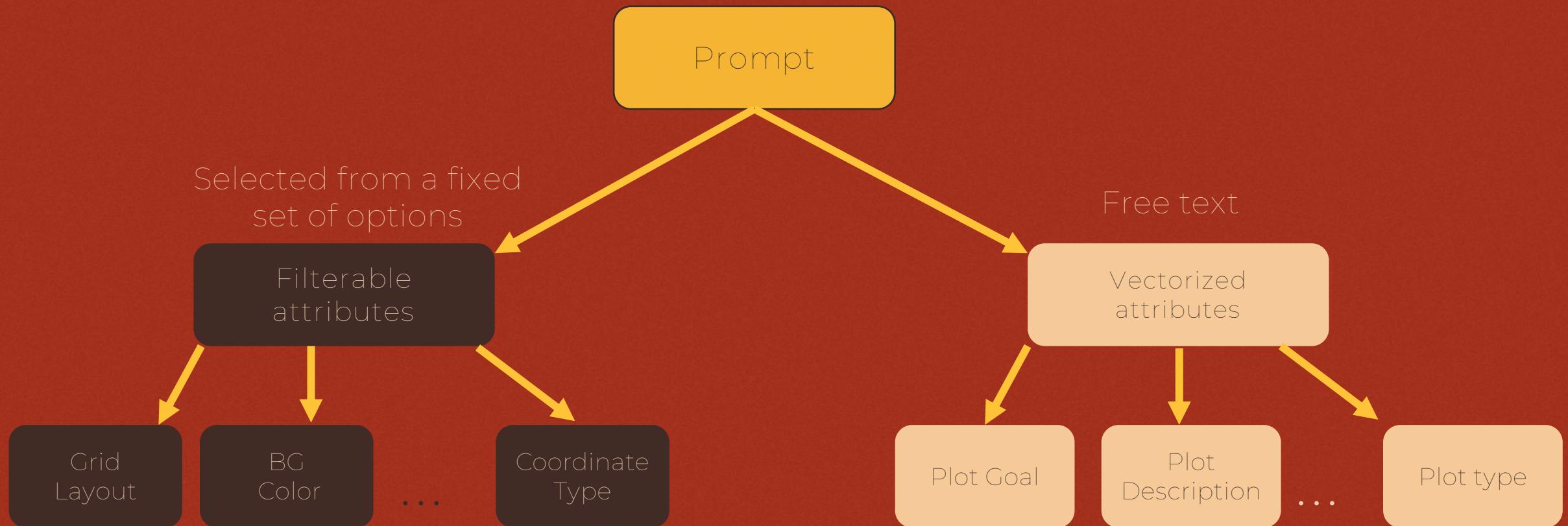
Some chart elements have only a limited set of possible values.



# Some chart elements have only a limited set of possible values.



# Deconstructing the Prompt



## Datawrapper



Plot  
Goals

Plot  
Description

BG  
Color

Coordinate  
Type

vectorized

vectorized

not vectorized

not vectorized



MISTRAL  
AI\_



# prompt\_even\_more\_structured\_final-v2-final-12-08-2025.txt

```
**Plot Type**
Primary_Category = (Use terms from the 30DayChartChallenge list where appropriate)
Subcategory = (Use terms from the 30DayChartChallenge list where appropriate)
Specific_Variant_Technique = (Free text)

**Theme & Grid**
Background_Color = (Simple color name)
Background_Type = "light" * "dark"
Grid_Color = (Simple color name)
Grid_Orientation = (MUST use one from: "horizontal", "vertical", "both", "radial", "none")
Grid_Solid = (MUST use one from: "single", "grid", "faceted", "small_multiples", "stacked", "side_by_side", "circular", "radial", "overlay", "irregular", "matrix")
Grid_Type = (MUST use one from: "major", "minor", "implicit", "subtle", "reference_lines", "none")
Grid_Style = (MUST use one or more from: "solid", "dashed", "dotted", "thin", "light", "subtle", "faint", "minimal", "none")

*Legend_Arrangement_Coordinates*
Legend_Position = (MUST use one from: "top_left", "top_center", "top_right", "bottom_left", "bottom_center", "bottom_right", "left", "right", "inline", "embedded", "inside", "none")
Legend_Orientation = (MUST use one from: "horizontal", "vertical", "inline", "embedded")
Arrangement_Type = (MUST use one or more from: "single", "grid", " facets", "small_multiples", "stacked", "side_by_side", "overlay", "hierarchical", "grouped", "clustered")
Coordinate_Type = (MUST use one from: "cartesian", "cartesian_3d", "polar", "circular", "geographic_general", "flow_network", "schematic", "linear", "categorical", "logarithmic", "mixed", "none")

**Typography**
Title_Font_Family = (Font name only)
Title_Font_Type = (MUST use one from: "sans-serif", "serif", "slab_serif", "script", "handwritten", "blackletter")
Title_Style = (MUST use one or more from: "regular", "bold", "italic", "semi-bold", "medium", "extra bold", "heavy", "mixed")
Subtitle_Font_Family = (Font name only)
Subtitle_Font_Type = (MUST use one or more from the list above)
Axis_Labels_Font_Family = (Font name only)
Axis_Labels_Font_Type = (MUST use one from the list above)
Axis_Labels_Style = (MUST use one or more from the list above)
Axis_Labels_Type = (MUST use one or more from the list above)

**Color**
Palette_Type = (MUST use one or more from: "sequential", "diverging", "qualitative", "categorical", "monochrome", "grayscale", "semantic", "brand colors", "highlight", "accent", "mixed palette")
Number_of_Distinct_Colors = (Integer only)

**Data_Sources**
Source_Name = (Free text)
Year = (Free text)

**Assessments**
Readability_Assessment = (MUST use one from: "very_high", "high", "medium_high", "medium", "medium_low", "low", "mixed", "none")
Readability_Assessment_Explanation = (One-sentence justification)
Color_Accessibility_Assessment = (MUST use one from: "excellent", "very_good", "good", "moderate", "fair", "poor", "not_applicable", "none")
Color_Accessibility_Assessment_Explanation = (One-sentence justification)
Text_Contrast_Ratio_WCAG = (MUST use one from: "aa_compliant", "aa_compliant", "aa_borderline", "below_aa", "not_assessed", "not_applicable", "none")
Text_Contrast_Ratio_WCAG_Explanation = (One-sentence justification)

XXX

Section_1_Data_and_Variable_Types@
Miles_per_Gallon: Continuous
Horsepower: Continuous
Weight: Continuous
Origin: Categorical
Model_Year: Temporal

XXX

Section_2_Variable_Mapping@
Horsepower is mapped to the x-axis via `aes(x = horsepower)`. Miles_per_Gallon is mapped to the y-axis via `aes(y = mpg)`. Origin is mapped to both color via `aes(color = origin)` and facet panels via `(-origin)`.

XXX

Section_3_Colour_Encoding_Details@
Key_Colours = Teal □#44A814*Green □#21918C*Yellow □#fde725
Semantic_Meaning = Color maps to country of origin using the `scale_color_viridis_d()` function.

XXX

Section_4_Chart_Element_Identification@
Statistical_Summaries = trend_lines*correlation_patterns

XXX

Section_5_Layout_Details@
Structure_Description = A 1x3 grid of scatter plots, arranged horizontally as defined by `facet_wrap` with `ncol=3`.
Aspect_Ratio = Each facet is approximately square.

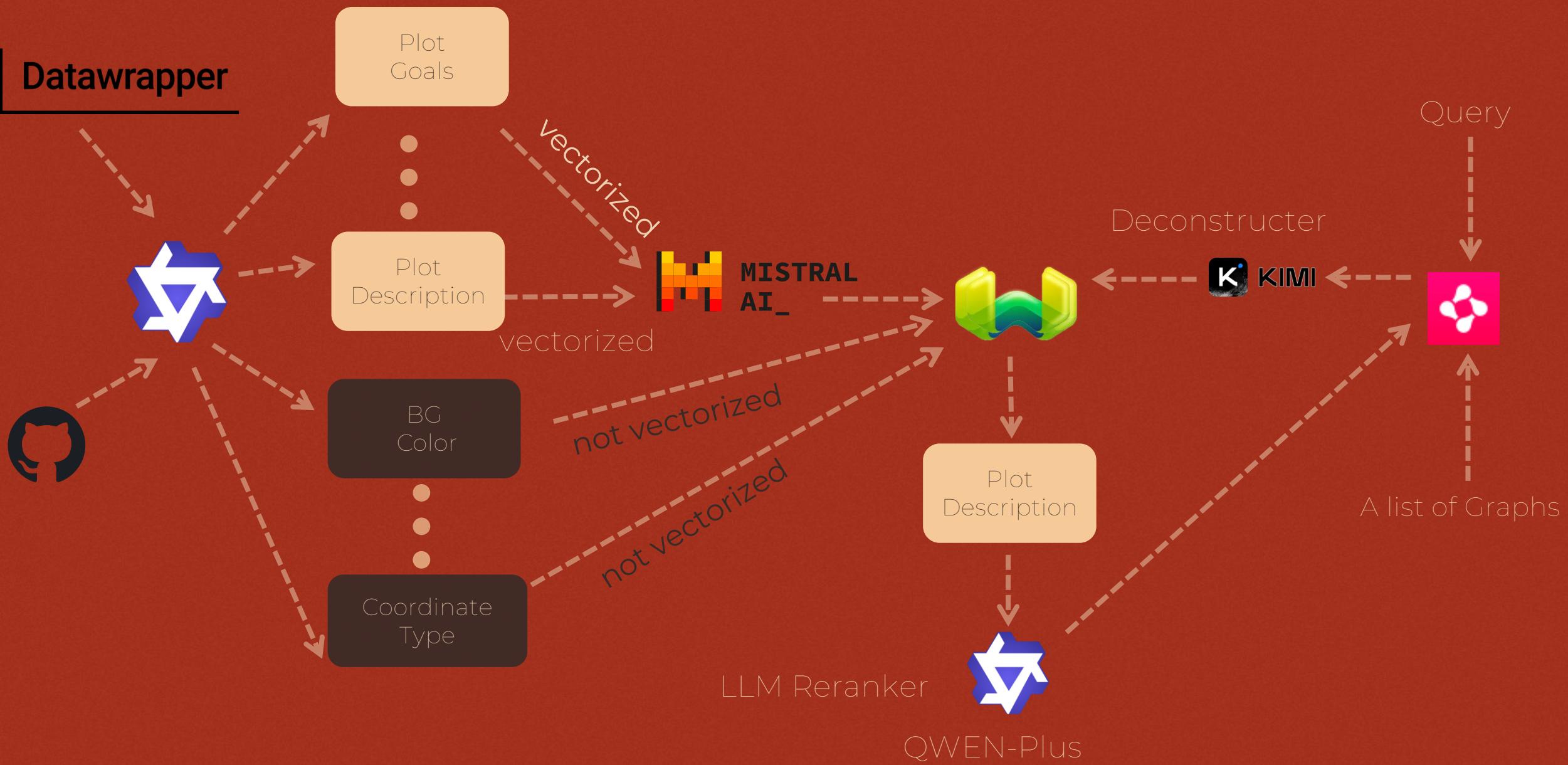
XXX

Section_6_Axes_and_Scales@
X_Axis_Horsepower:
Scale_Type = Linear
Position = Bottom
Y_Axis_Miles_per_Gallon:
Scale_Type = Linear
Position = Left

XXX

Section_7_annotation_and_storytelling_elements@
Annotations = trend_annotations*contextual_info
```

## Datawrapper



# **MODE 3**

## **USING LONGER CONTEXT**

# RAG IS DEAD!

The screenshot shows a video player interface. On the left is a thumbnail for a video titled "Long Live Context Engineering - with Jeff Huber of Chroma". The thumbnail features a man smiling, the Chroma logo, and the text "CO-FOUNDER" above "RAG IS DEAD". The video duration is 57:01. On the right, the main video frame displays the title "Long Live Context Engineering - with Jeff Huber of Chroma" and the subtitle "Jeff Huber of Chroma joins us to talk about what actually matters in vector databases in 2025, why "modern search for AI" is ...". Below the video frame, there is a "Matching chapter" section with a thumbnail for "12:15 Context Engineering and the Problems with RAG".

Long Live Context Engineering - with Jeff Huber of Chroma

46K views • 1 month ago

Latent Space

Jeff Huber of Chroma joins us to talk about what actually matters in vector databases in 2025, why "modern search for AI" is ...

Matching chapter 12:15 Context Engineering and the Problems with RAG

# Datawrapper



QWEN-VL-Max



MISTRAL  
AI\_



Query Rewriter



Query

A list of Graphs



Plot  
Description



LLM Reranker

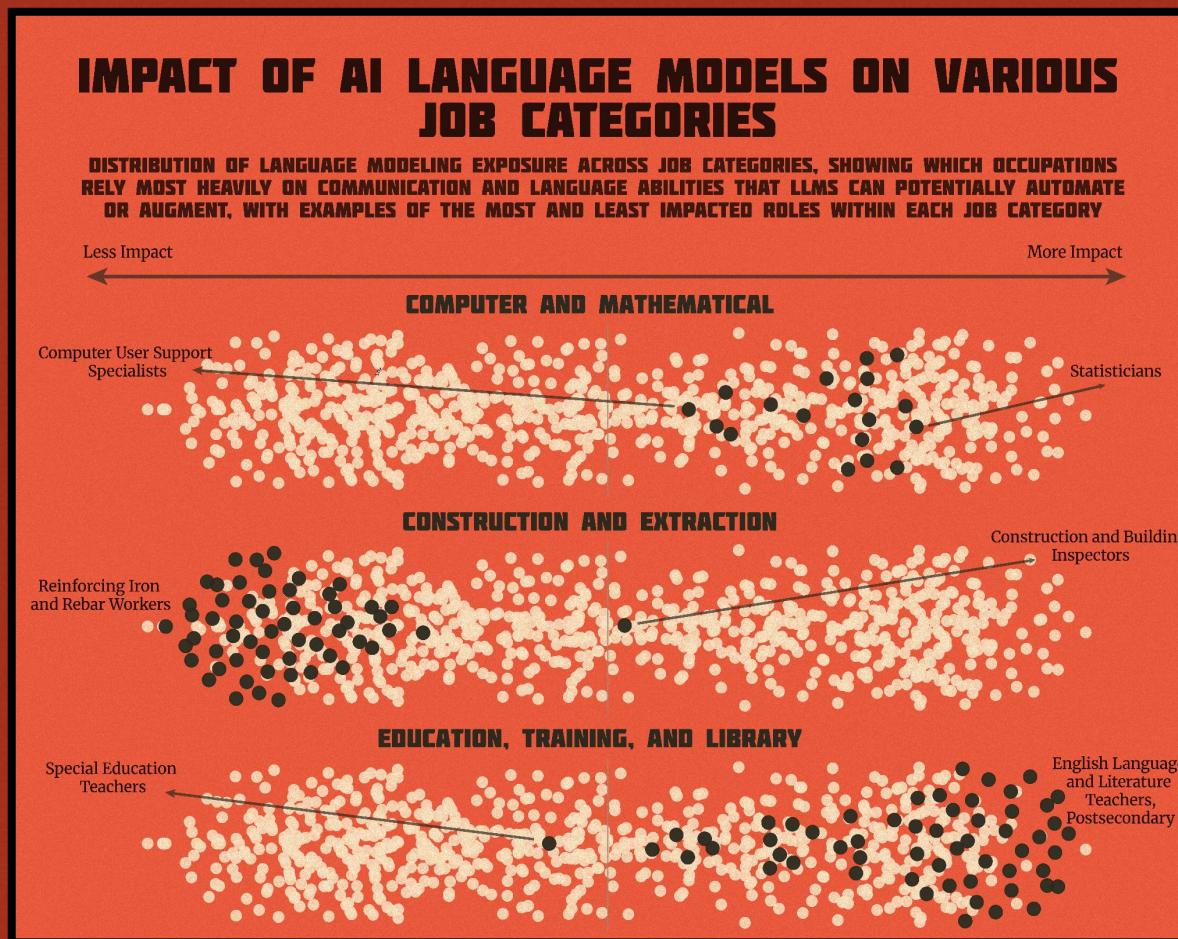
QWEN-Plus



# LIMITATIONS

# Garbage In, Garbage Out

# The Challenge of Subjectivity



Knowing What You Didn't Find

# Evaluation is Hard

# **IMPROVEMENTS**

Better prompt, better VLM model

# Better toolcalling logic

# Better query rewriting

# **TAKEAWAY**

# TAKEAWAY

*“Stop using your head as an open browser tab, instead use RAGs”*

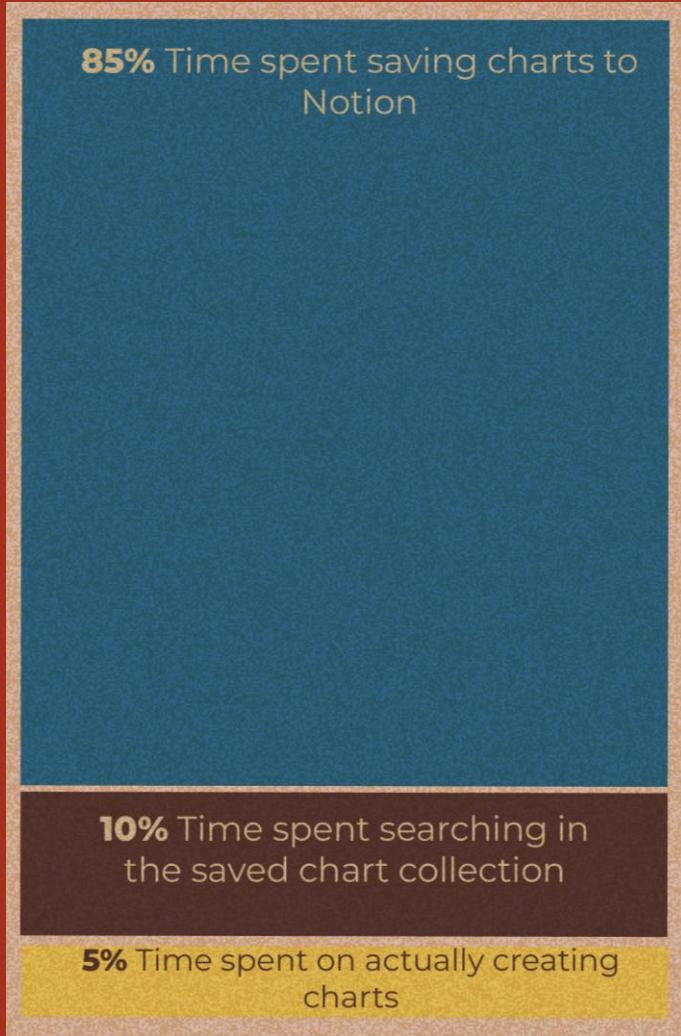


**WHAT IS NEXT ...**

# How it started



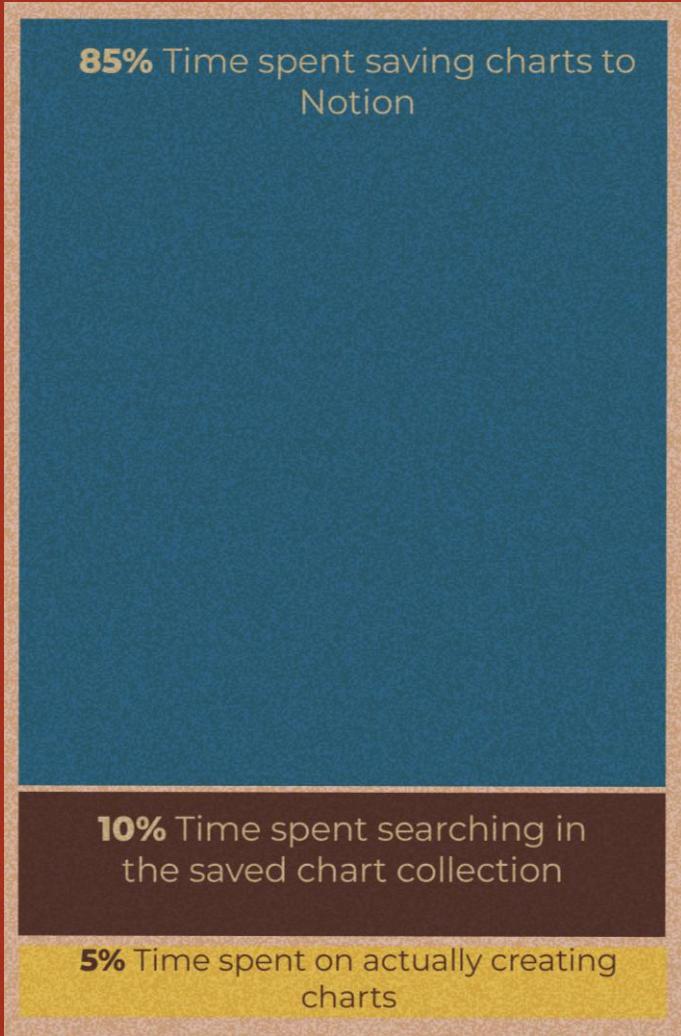
## How it started



## How I expected it to go



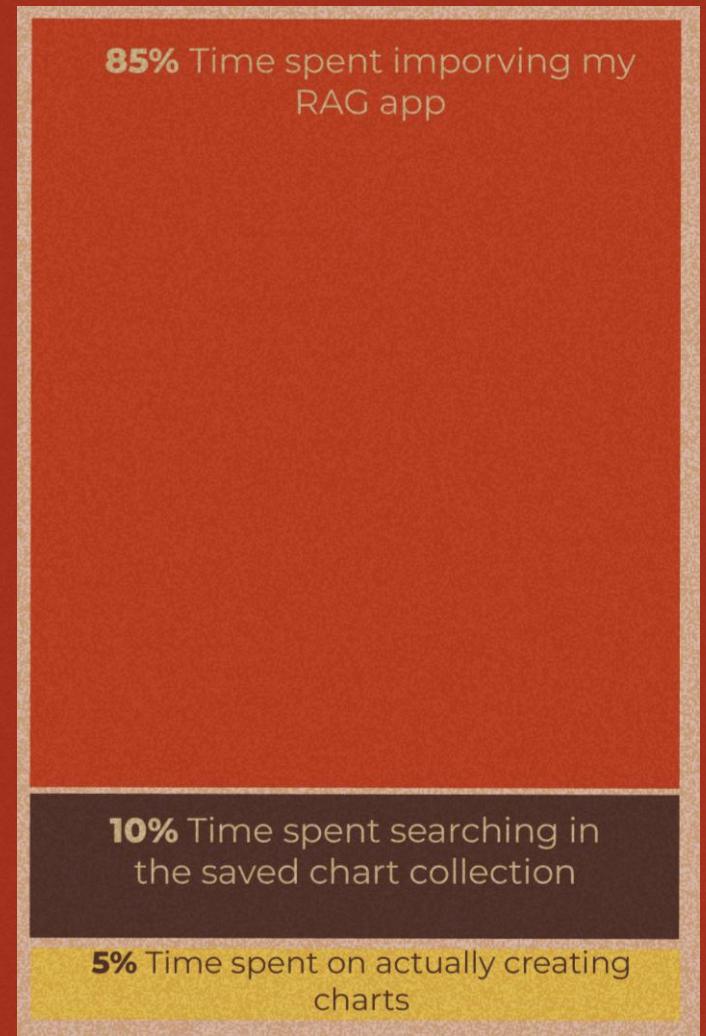
## How it started



## How I expected it to go



## How it's going



**THANK YOU**

# **APPENDIX**

# Using AI in Creative work



Never AI camp

Vibe Artists

Audrey's Weekly Email Newsletter

## Why I'm betting against AI.

Is AI the future? I don't fucking give a shit.

AUDREY KNOX  
AUG 15, 2025

Share

162 94 52

This is a free edition of my Weekly Email Newsletter! Every week I do a deep dive on an aspect of the craft and business of screenwriting from a (former) literary manager's perspective.

Subscribe, so you never miss a post.

mcnakhaee@gmail.com Subscribe

Discourse on whether AI is ethical and what the future looks like for it (and for us) has become unavoidable.

I've procrastinated writing about it because my opinion on the topic doesn't feel necessary to add to the din.

Why I'm betting against AI.

## The success of AI music creators sparks a debate on the future of the music industry

People complain "that you're using a computer to do all the work for you. I don't see it that way. I see it as any other tool that we have," said Scott Smith, whose AI band, Pulse Empire, was inspired by 1980s British synthesizer-driven groups like New Order and Depeche Mode.

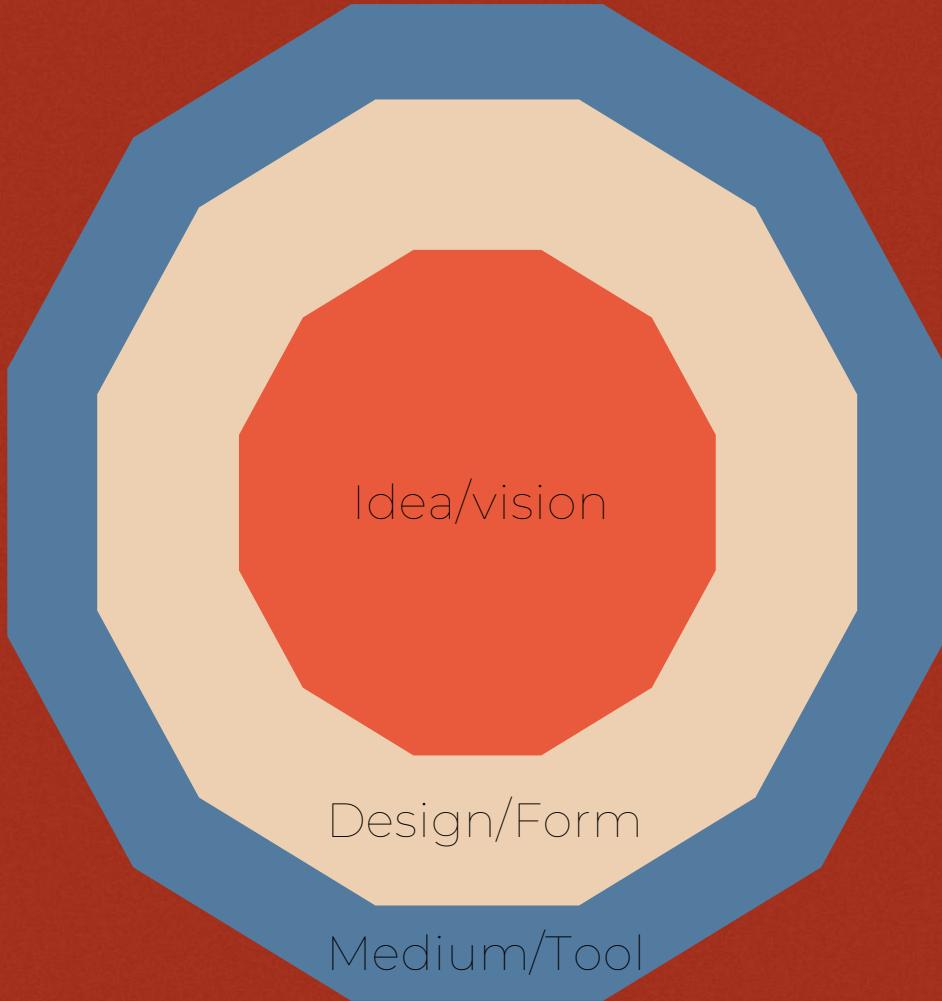
Smith, 56 and a semi-retired former U.S. Navy public affairs officer in Portland, Oregon, said "music producers have lots of tools in their arsenal" to enhance recordings that listeners aren't aware of.

Like McCann, Smith never mastered a musical instrument. Both say they put lots of time and effort into crafting their music.

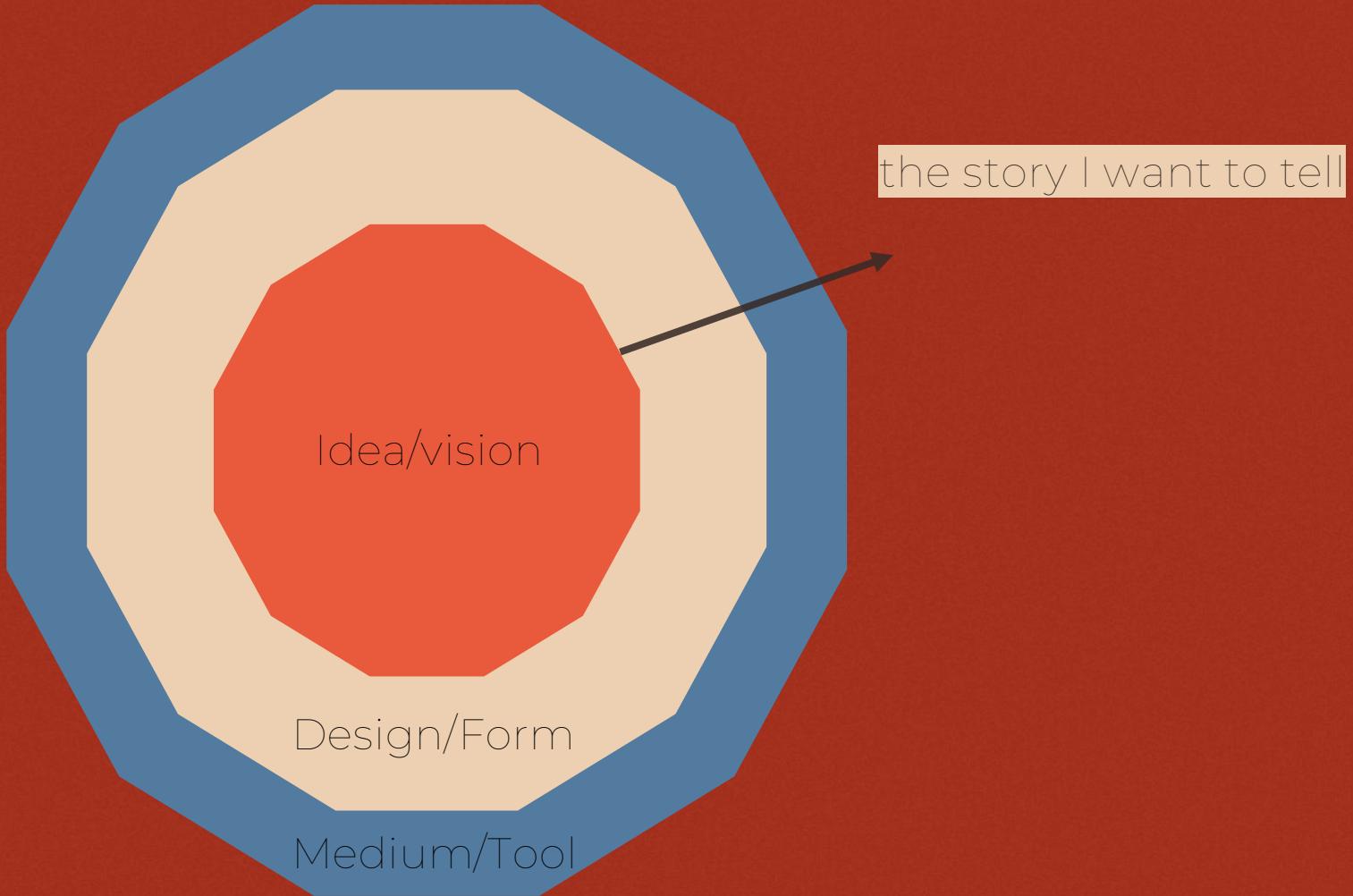
Once Smith gets inspiration, it takes him just 10 minutes to write the lyrics. But then he'll spend as much as eight to nine hours generating different versions until the song "matches my vision."

The success of AI music creators sparks a debate on the future of the music industry

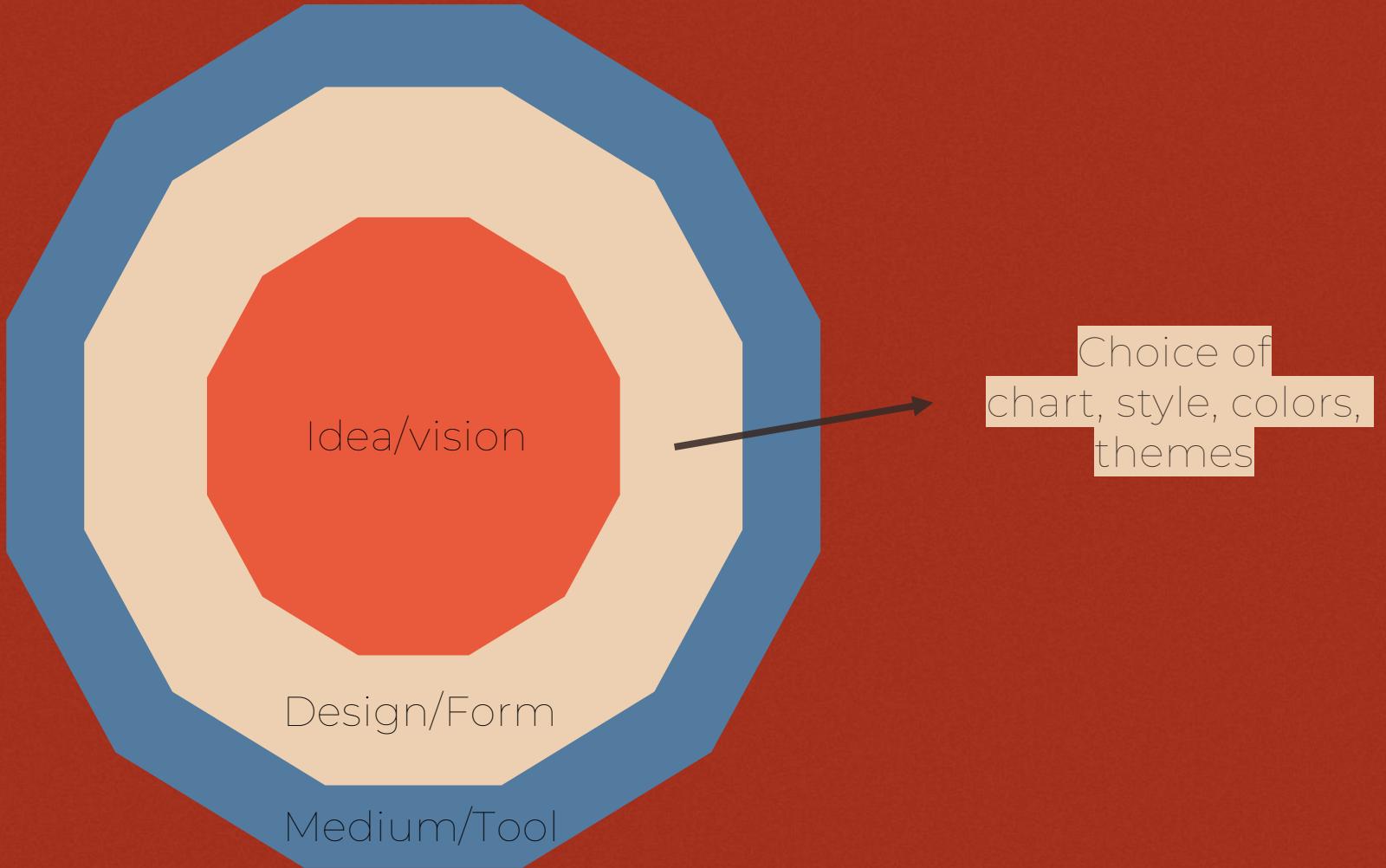
# Creativity Blob



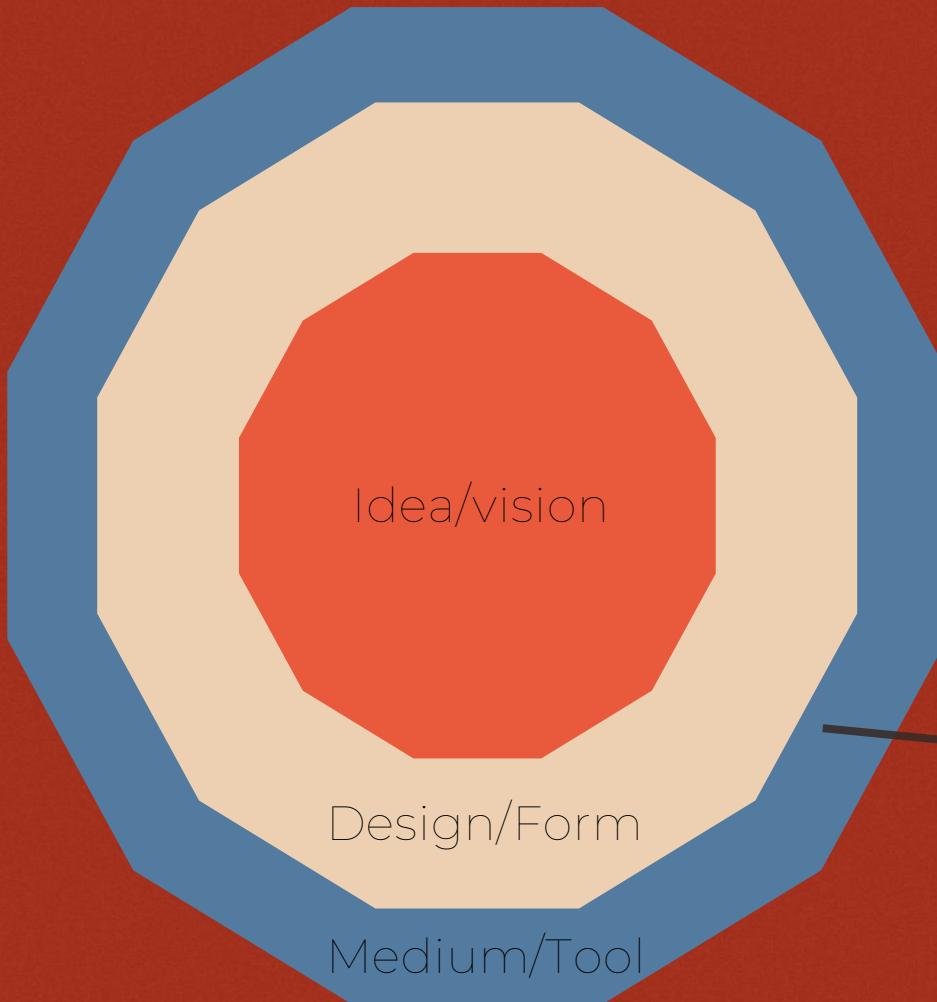
# Creativity Blob



# Creativity Blob

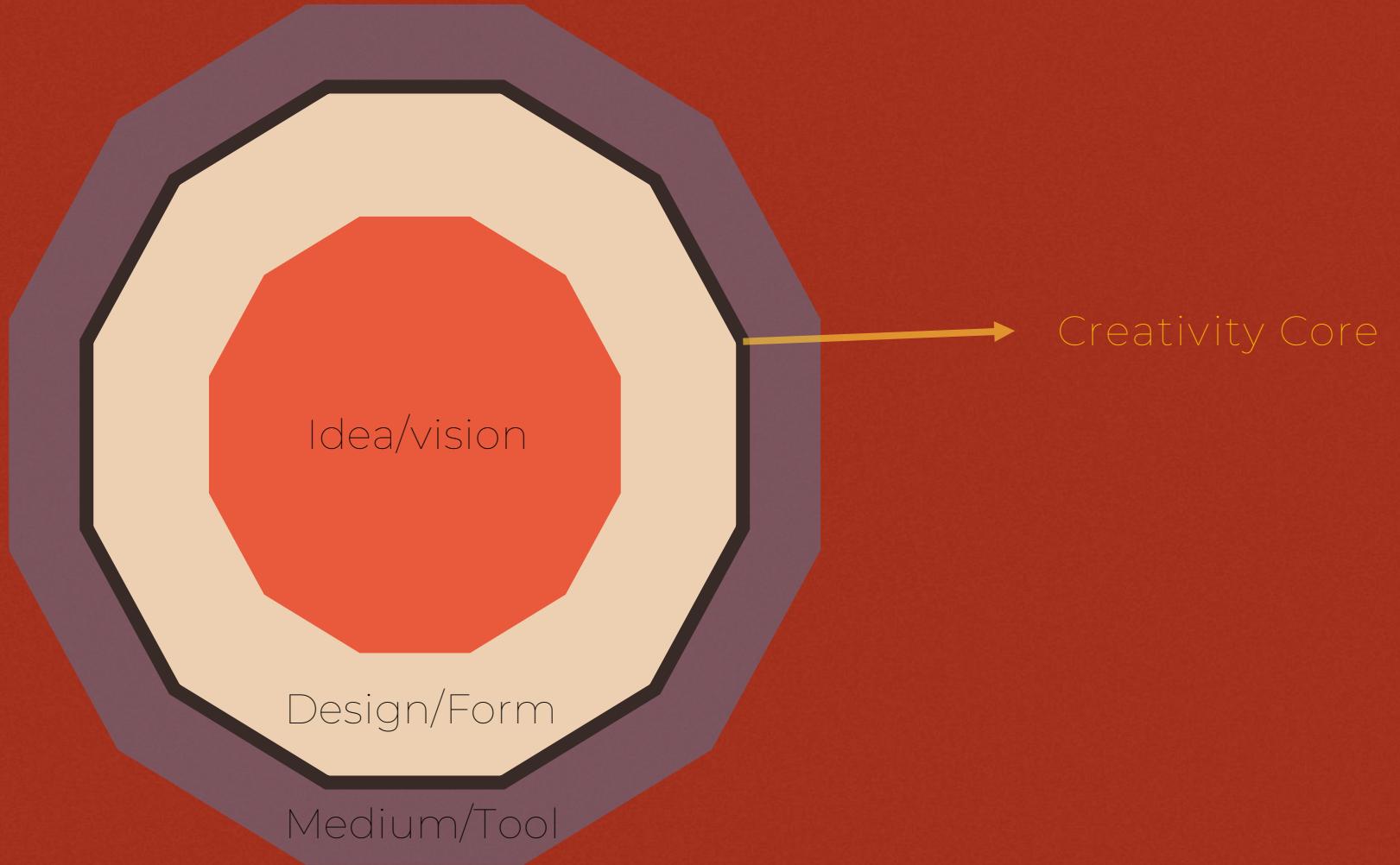


# My Rule

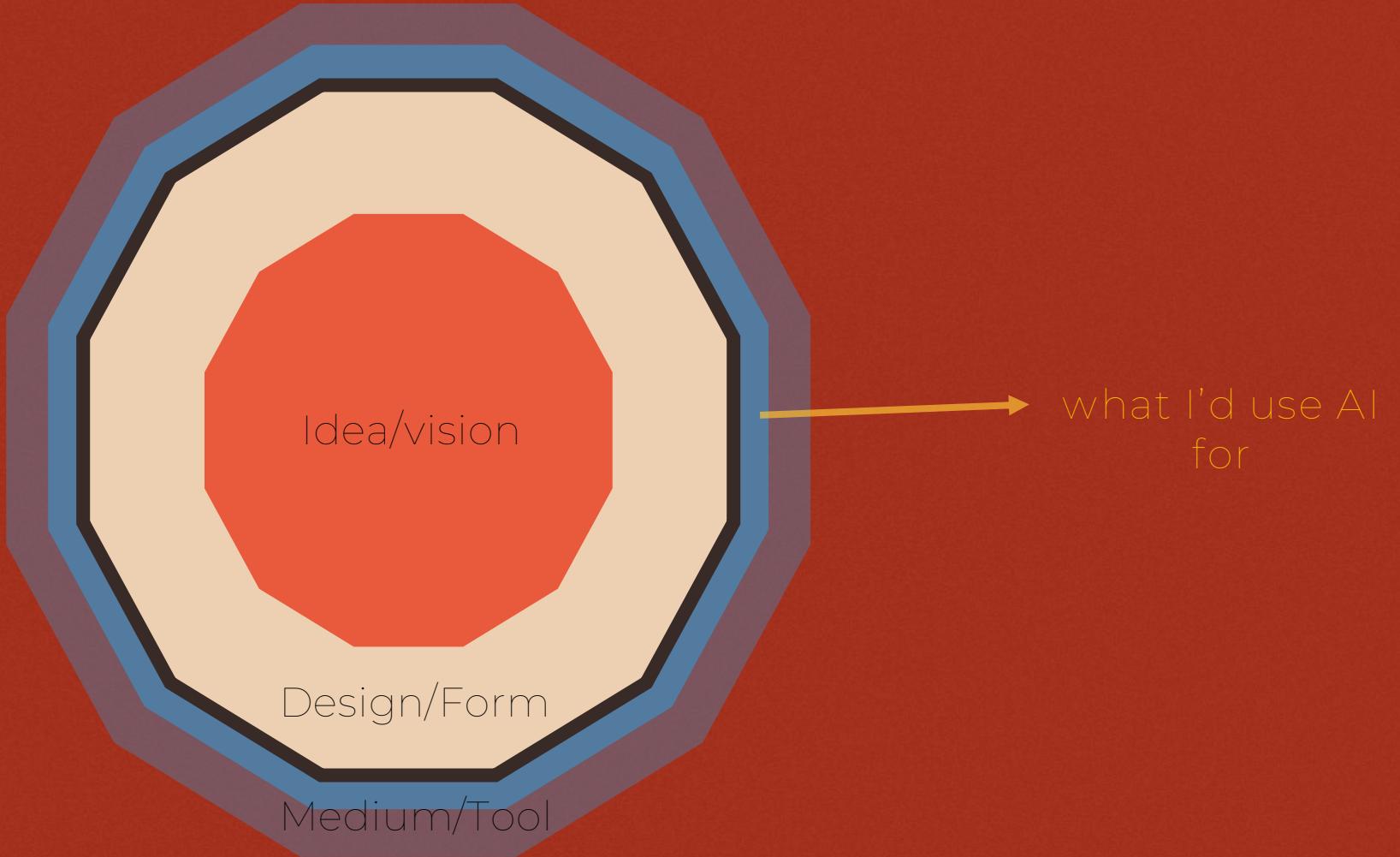


ggplot2, Illustrator,  
plotnine, etc.

# Creativity Blob



# Creativity Blob



# My Rule-of-Thumb for using AI

## Where I'd use AI

For my data research and exploring data sources

To write the entire data scraping code

To proofread my title, subtitles, captions, etc.

## Where I draw a line in using AI

To give me inspirations, ideas or brainstorm with me

To decide which variables to visualize

To recommend me chart types.

To recommend or pick themes, colors, and overall aesthetics for me

To generate the entire ggplot2 code

# Data

Each raw/point = occupation

Variables:

- Job Category
- Impact Score
- Job Title

# IMPACT OF AI LANGUAGE MODELS ON VARIOUS JOB CATEGORIES

DISTRIBUTION OF LANGUAGE MODELING EXPOSURE ACROSS JOB CATEGORIES, SHOWING WHICH OCCUPATIONS RELY MOST HEAVILY ON COMMUNICATION AND LANGUAGE ABILITIES THAT LLMS CAN POTENTIALLY AUTOMATE OR AUGMENT, WITH EXAMPLES OF THE MOST AND LEAST IMPACTED ROLES WITHIN EACH JOB CATEGORY

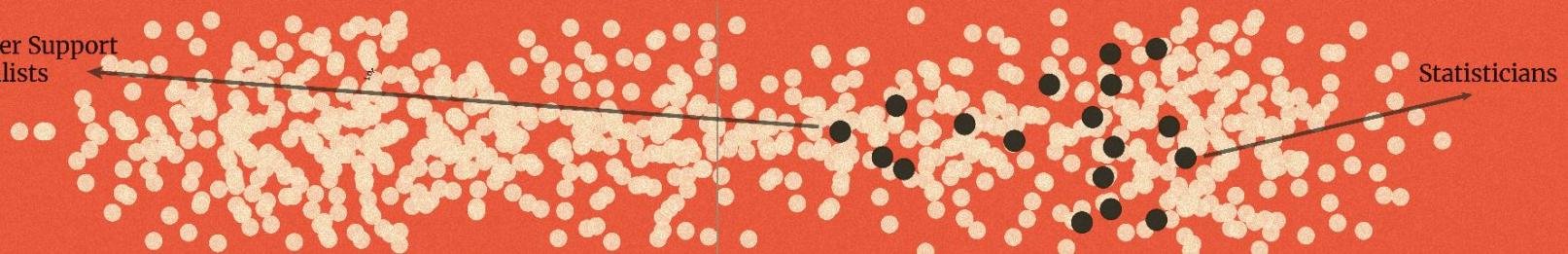
Less Impact

More Impact

## COMPUTER AND MATHEMATICAL

Computer User Support Specialists

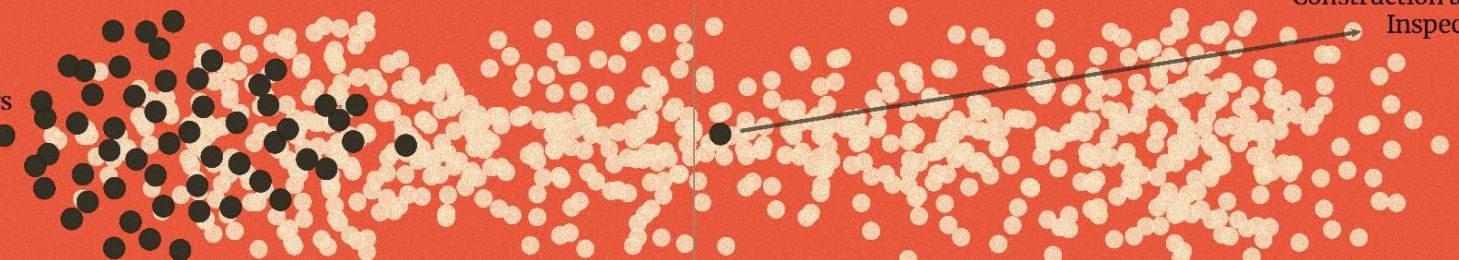
Statisticians



## CONSTRUCTION AND EXTRACTION

Reinforcing Iron and Rebar Workers

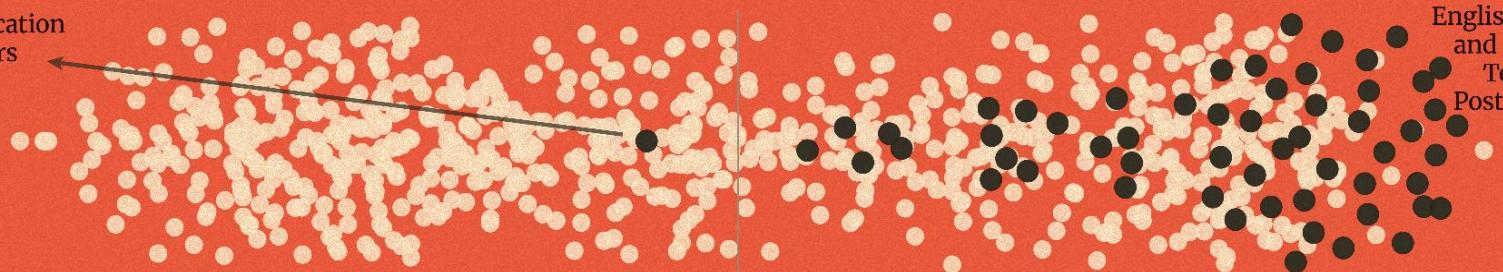
Construction and Building Inspectors



## EDUCATION, TRAINING, AND LIBRARY

Special Education Teachers

English Language and Literature Teachers, Postsecondary



# Aesthetics

x-axis → Impact score

y-axis → random position

color/fill → Two groups:

highlighted jobs (black) vs.

general jobs (white).

## IMPACT OF AI LANGUAGE MODELS ON VARIOUS JOB CATEGORIES

DISTRIBUTION OF LANGUAGE MODELING EXPOSURE ACROSS JOB CATEGORIES, SHOWING WHICH OCCUPATIONS RELY MOST HEAVILY ON COMMUNICATION AND LANGUAGE ABILITIES THAT LLMs CAN POTENTIALLY AUTOMATE OR AUGMENT, WITH EXAMPLES OF THE MOST AND LEAST IMPACTED ROLES WITHIN EACH JOB CATEGORY

Less Impact

More Impact

### COMPUTER AND MATHEMATICAL

Computer User Support Specialists

Statisticians

### CONSTRUCTION AND EXTRACTION

Reinforcing Iron and Rebar Workers

Construction and Building Inspectors

### EDUCATION, TRAINING, AND LIBRARY

Special Education Teachers

English Language and Literature Teachers, Postsecondary

# Geometries

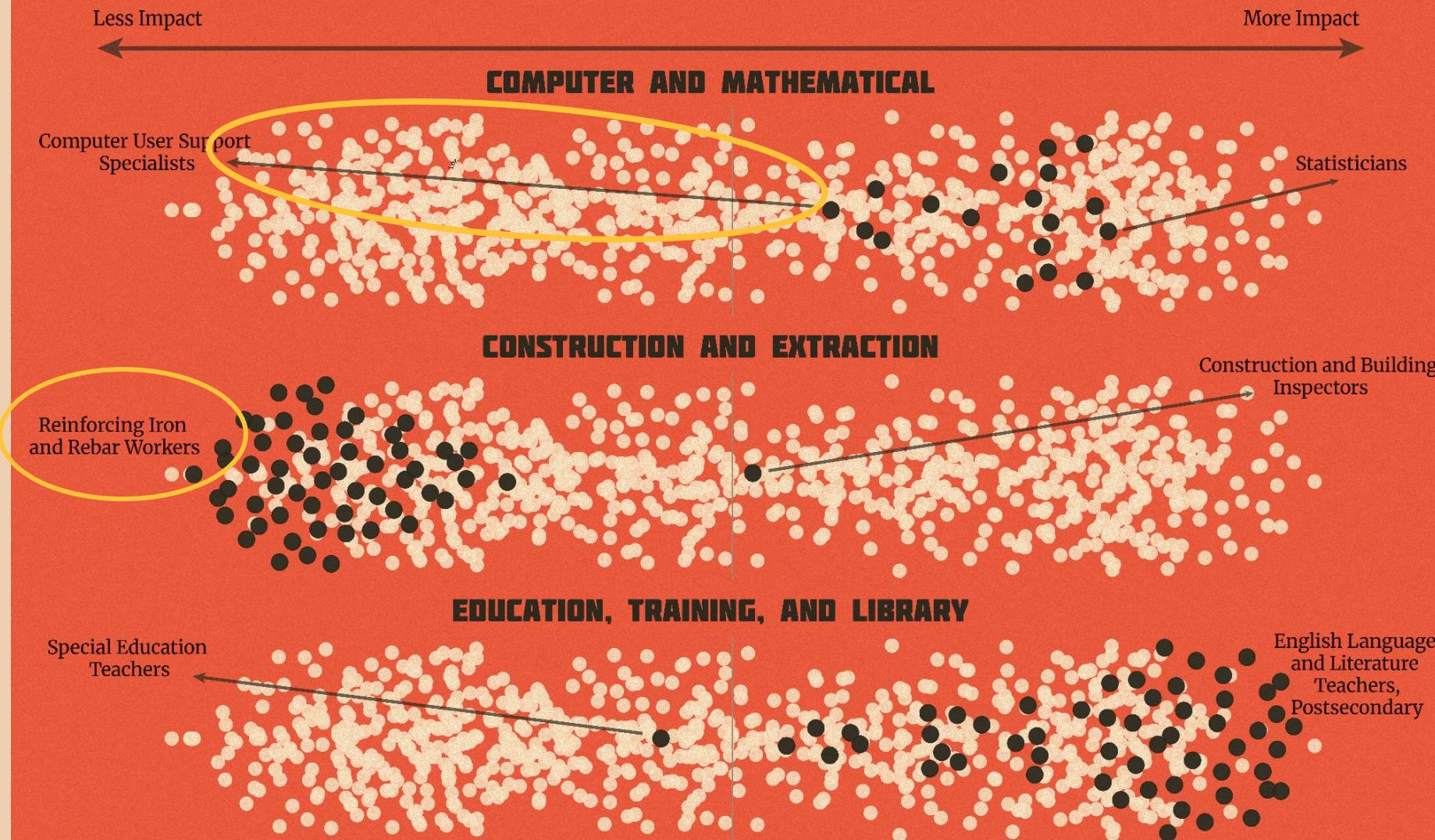
`geom_point()` → For scatter of jobs within categories.

`geom_text()` → For labeling specific jobs.

`geom_segment()` → For arrows connecting labels to points.

# IMPACT OF AI LANGUAGE MODELS ON VARIOUS JOB CATEGORIES

DISTRIBUTION OF LANGUAGE MODELING EXPOSURE ACROSS JOB CATEGORIES, SHOWING WHICH OCCUPATIONS RELY MOST HEAVILY ON COMMUNICATION AND LANGUAGE ABILITIES THAT LLMS CAN POTENTIALLY AUTOMATE OR AUGMENT, WITH EXAMPLES OF THE MOST AND LEAST IMPACTED ROLES WITHIN EACH JOB CATEGORY



# Scales

x scale: Continuous

color scale: Categorical

(highlighted = **black**,

non-highlighted = **white**)

# IMPACT OF AI LANGUAGE MODELS ON VARIOUS JOB CATEGORIES

DISTRIBUTION OF LANGUAGE MODELING EXPOSURE ACROSS JOB CATEGORIES, SHOWING WHICH OCCUPATIONS RELY MOST HEAVILY ON COMMUNICATION AND LANGUAGE ABILITIES THAT LLMS CAN POTENTIALLY AUTOMATE OR AUGMENT, WITH EXAMPLES OF THE MOST AND LEAST IMPACTED ROLES WITHIN EACH JOB CATEGORY

Less Impact

More Impact

## COMPUTER AND MATHEMATICAL

Computer User Support Specialists

Statisticians

## CONSTRUCTION AND EXTRACTION

Reinforcing Iron and Rebar Workers

Construction and Building Inspectors

## EDUCATION, TRAINING, AND LIBRARY

Special Education Teachers

English Language and Literature Teachers, Postsecondary

# Facets

3 major facets/panels

## IMPACT OF AI LANGUAGE MODELS ON VARIOUS JOB CATEGORIES

DISTRIBUTION OF LANGUAGE MODELING EXPOSURE ACROSS JOB CATEGORIES, SHOWING WHICH OCCUPATIONS RELY MOST HEAVILY ON COMMUNICATION AND LANGUAGE ABILITIES THAT LLMS CAN POTENTIALLY AUTOMATE OR AUGMENT, WITH EXAMPLES OF THE MOST AND LEAST IMPACTED ROLES WITHIN EACH JOB CATEGORY

Less Impact

More Impact

### COMPUTER AND MATHEMATICAL

Computer User Support Specialists

Statisticians

### CONSTRUCTION AND EXTRACTION

Reinforcing Iron and Rebar Workers

Construction and Building Inspectors

### EDUCATION, TRAINING, AND LIBRARY

Special Education Teachers

English Language and Literature Teachers, Postsecondary

# Coordinate

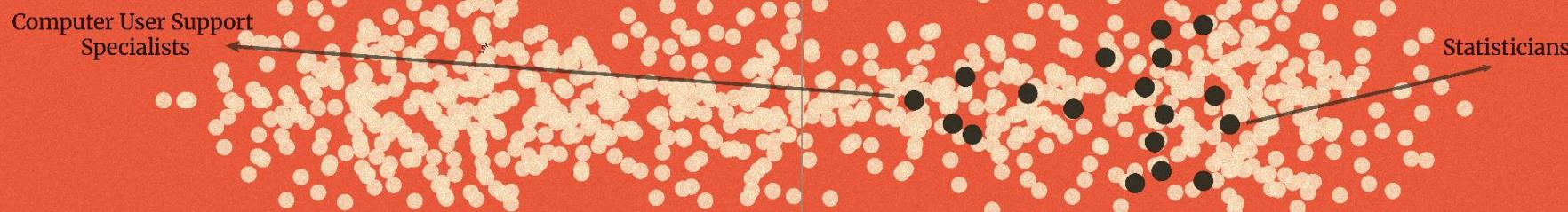
`coord_cartesian()`: standard  
Cartesian coordinates

# IMPACT OF AI LANGUAGE MODELS ON VARIOUS JOB CATEGORIES

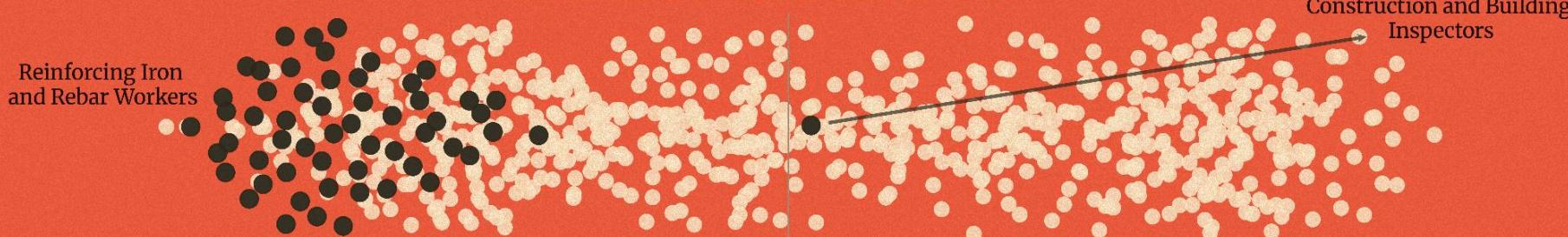
DISTRIBUTION OF LANGUAGE MODELING EXPOSURE ACROSS JOB CATEGORIES, SHOWING WHICH OCCUPATIONS RELY MOST HEAVILY ON COMMUNICATION AND LANGUAGE ABILITIES THAT LLMS CAN POTENTIALLY AUTOMATE OR AUGMENT, WITH EXAMPLES OF THE MOST AND LEAST IMPACTED ROLES WITHIN EACH JOB CATEGORY

Less Impact ← → More Impact

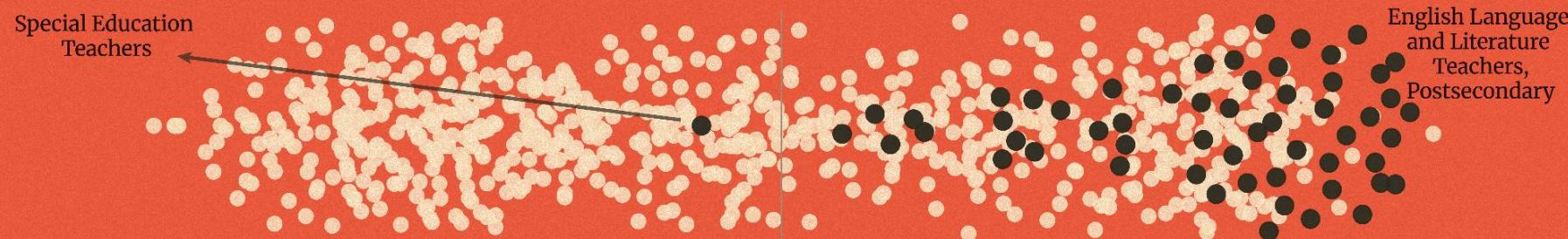
## COMPUTER AND MATHEMATICAL



## CONSTRUCTION AND EXTRACTION



## EDUCATION, TRAINING, AND LIBRARY



# Theme

Background: #E95A3C

Font family (**MOLOT**,

Merriweather)

Font color (black),

Font style (**bold**)

# IMPACT OF AI LANGUAGE MODELS ON VARIOUS JOB CATEGORIES

DISTRIBUTION OF LANGUAGE MODELING EXPOSURE ACROSS JOB CATEGORIES, SHOWING WHICH OCCUPATIONS RELY MOST HEAVILY ON COMMUNICATION AND LANGUAGE ABILITIES THAT LLMS CAN POTENTIALLY AUTOMATE OR AUGMENT, WITH EXAMPLES OF THE MOST AND LEAST IMPACTED ROLES WITHIN EACH JOB CATEGORY

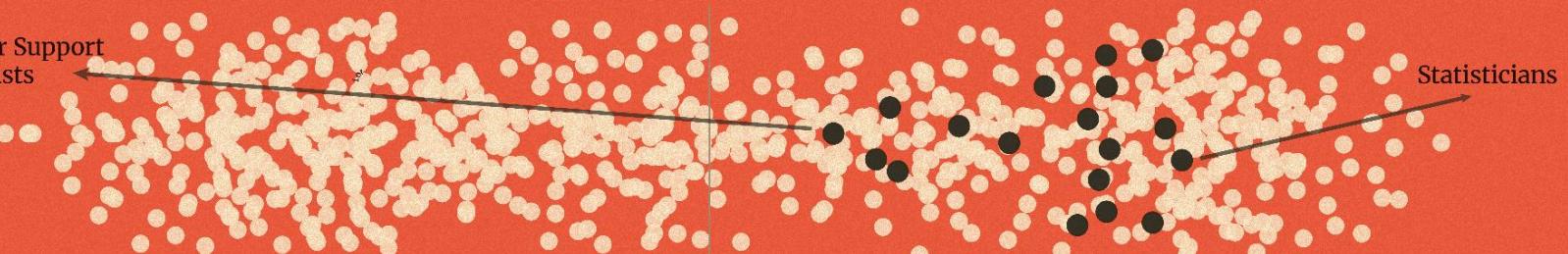
Less Impact

More Impact

## COMPUTER AND MATHEMATICAL

Computer User Support Specialists

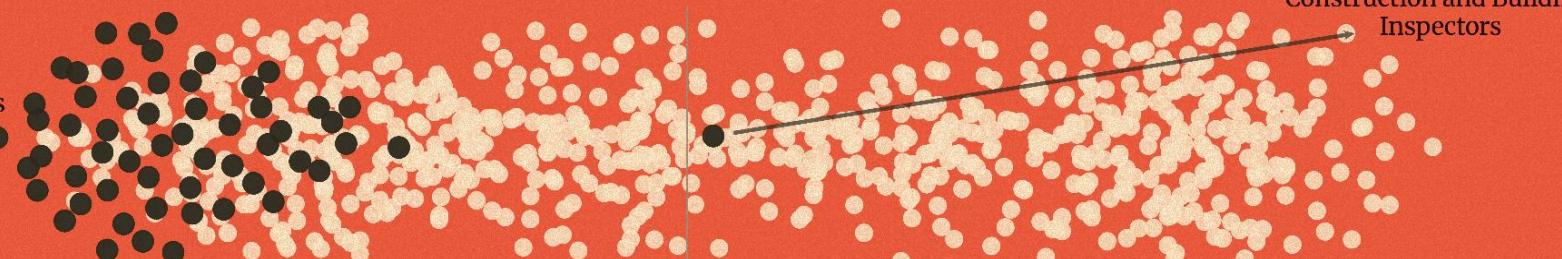
Statisticians



## CONSTRUCTION AND EXTRACTION

Reinforcing Iron and Rebar Workers

Construction and Building Inspectors



## EDUCATION, TRAINING, AND LIBRARY

Special Education Teachers

English Language and Literature Teachers, Postsecondary



# Scales

x scale: Continuous

color scale: Categorical

(highlighted = **black**,

non-highlighted = **white**)

# IMPACT OF AI LANGUAGE MODELS ON VARIOUS JOB CATEGORIES

DISTRIBUTION OF LANGUAGE MODELING EXPOSURE ACROSS JOB CATEGORIES, SHOWING WHICH OCCUPATIONS RELY MOST HEAVILY ON COMMUNICATION AND LANGUAGE ABILITIES THAT LLMS CAN POTENTIALLY AUTOMATE OR AUGMENT, WITH EXAMPLES OF THE MOST AND LEAST IMPACTED ROLES WITHIN EACH JOB CATEGORY

Less Impact

More Impact

## COMPUTER AND MATHEMATICAL

Computer User Support Specialists

Statisticians

## CONSTRUCTION AND EXTRACTION

Reinforcing Iron and Rebar Workers

Construction and Building Inspectors

## EDUCATION, TRAINING, AND LIBRARY

Special Education Teachers

English Language and Literature Teachers, Postsecondary

# Aesthetics

x-axis → Impact score

y-axis → random position

generated by the beeswarm  
algorithm

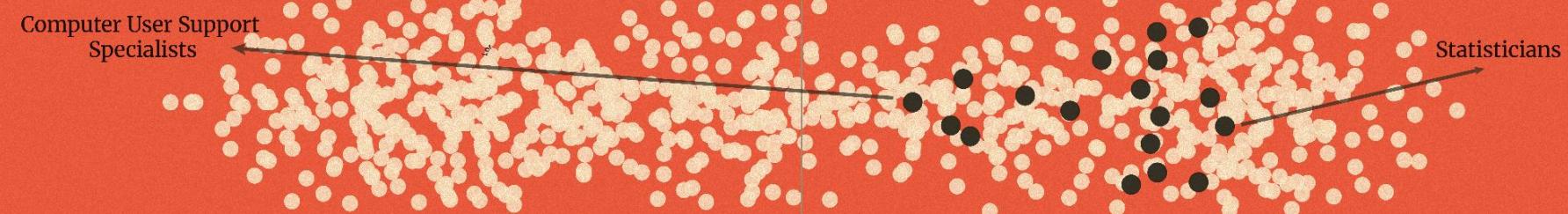
color/fill → Two groups:  
highlighted jobs (**black**) vs.  
general jobs (**white**).

## IMPACT OF AI LANGUAGE MODELS ON VARIOUS JOB CATEGORIES

DISTRIBUTION OF LANGUAGE MODELING EXPOSURE ACROSS JOB CATEGORIES, SHOWING WHICH OCCUPATIONS RELY MOST HEAVILY ON COMMUNICATION AND LANGUAGE ABILITIES THAT LLMs CAN POTENTIALLY AUTOMATE OR AUGMENT, WITH EXAMPLES OF THE MOST AND LEAST IMPACTED ROLES WITHIN EACH JOB CATEGORY

Less Impact ← More Impact →

### COMPUTER AND MATHEMATICAL



### CONSTRUCTION AND EXTRACTION



### EDUCATION, TRAINING, AND LIBRARY

