

Data 227

Autumn, 2023

Encodings



STYLIZED BY SPENCER HILL

10 Leading Causes of Death by Age Group, United States – 2010

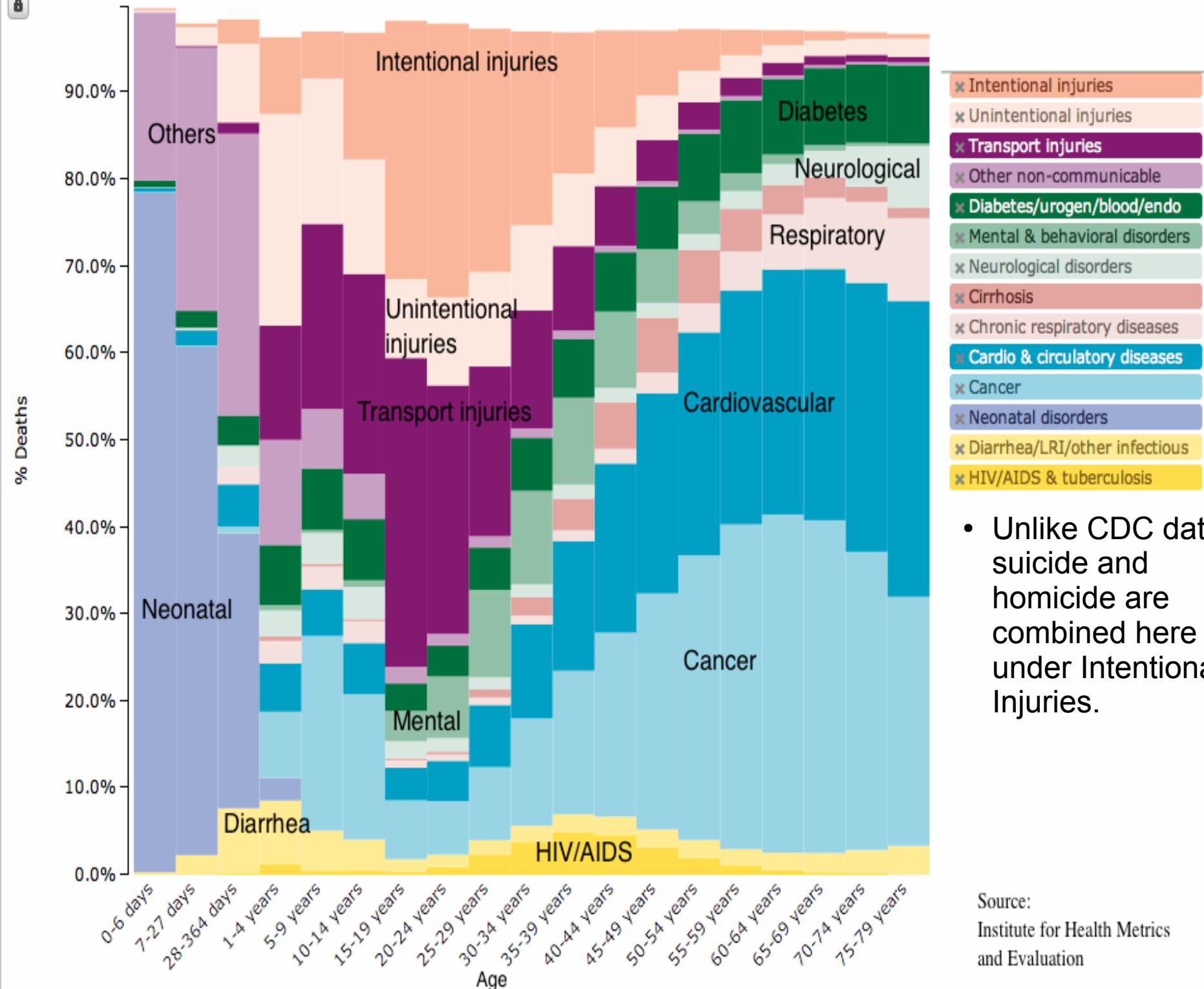
Rank	Age Groups											Total
	<1	1-4	5-9	10-14	15-24	25-34	35-44	45-54	55-64	65+		
1	Congenital Anomalies 5,107	Unintentional Injury 1,394	Unintentional Injury 758	Unintentional Injury 885	Unintentional Injury 12,341	Unintentional Injury 14,573	Unintentional Injury 14,792	Malignant Neoplasms 50,211	Malignant Neoplasms 109,501	Heart Disease 477,338	Heart Disease 597,689	
2	Short Gestation 4,148	Congenital Anomalies 507	Malignant Neoplasms 439	Malignant Neoplasms 477	Homicide 4,678	Suicide 5,735	Malignant Neoplasms 11,809	Heart Disease 36,729	Heart Disease 68,077	Malignant Neoplasms 396,670	Malignant Neoplasms 574,743	
3	SIDS 2,063	Homicide 385	Congenital Anomalies 163	Suicide 267	Suicide 4,600	Homicide 4,258	Heart Disease 10,594	Unintentional Injury 19,667	Chronic Low Respiratory Disease 14,242	Chronic Low Respiratory Disease 118,031	Chronic Low Respiratory Disease 138,080	
4	Maternal Pregnancy Comp. 1,561	Malignant Neoplasms 346	Homicide 111	Homicide 150	Malignant Neoplasms 1,604	Malignant Neoplasms 3,619	Suicide 6,571	Suicide 8,799	Unintentional Injury 14,023	Cerebro-vascular 109,990	Cerebro-vascular 129,476	
5	Unintentional Injury 1,110	Heart Disease 159	Heart Disease 68	Congenital Anomalies 135	Heart Disease 1,028	Heart Disease 3,222	Homicide 2,473	Liver Disease 8,651	Diabetes Mellitus 11,677	Alzheimer's Disease 82,616	Unintentional Injury 120,859	
6	Placenta Cord. Membranes 1,030	Influenza & Pneumonia 91	Chronic Low Respiratory Disease 60	Heart Disease 117	Congenital Anomalies 412	HIV 741	Liver Disease 2,423	Cerebro-vascular 5,910	Cerebro-vascular 10,693	Diabetes Mellitus 49,191	Alzheimer's Disease 83,494	
7	Bacterial Sepsis 583	Septicemia 62	Cerebro-vascular 47	Chronic Low Respiratory Disease 73	Cerebro-vascular 190	Diabetes Mellitus 606	Cerebro-vascular 1,904	Diabetes Mellitus 5,610	Liver Disease 9,764	Influenza & Pneumonia 42,846	Diabetes Mellitus 69,071	
8	Respiratory Distress 514	Benign Neoplasms 59	Benign Neoplasms 37	Benign Neoplasms 45	Influenza & Pneumonia 181	Cerebro-vascular 517	HIV 1,898	Chronic Low Respiratory Disease 4,452	Suicide 6,384	Nephritis 41,994	Nephritis 50,476	
9	Circulatory System Disease 507	Perinatal Period 52	Influenza & Pneumonia 37	Cerebro-vascular 43	Diabetes Mellitus 165	Liver Disease 487	Diabetes Mellitus 1,789	HIV 3,123	Nephritis 5,082	Unintentional Injury 41,300	Influenza & Pneumonia 50,097	
10	Necrotizing Enterocolitis 472	Chronic Low Respiratory Disease 51	Septicemia 32	Septicemia 35	Complicated Pregnancy 163	Congenital Anomalies 397	Influenza & Pneumonia 773	Viral Hepatitis 2,376	Septicemia 4,604	Septicemia 26,310	Suicide 38,364	

Data Source: National Vital Statistics System, National Center for Health Statistics, CDC.

Produced by: Office of Statistics and Programming, National Center for Injury Prevention and Control, CDC using WISQARS™.

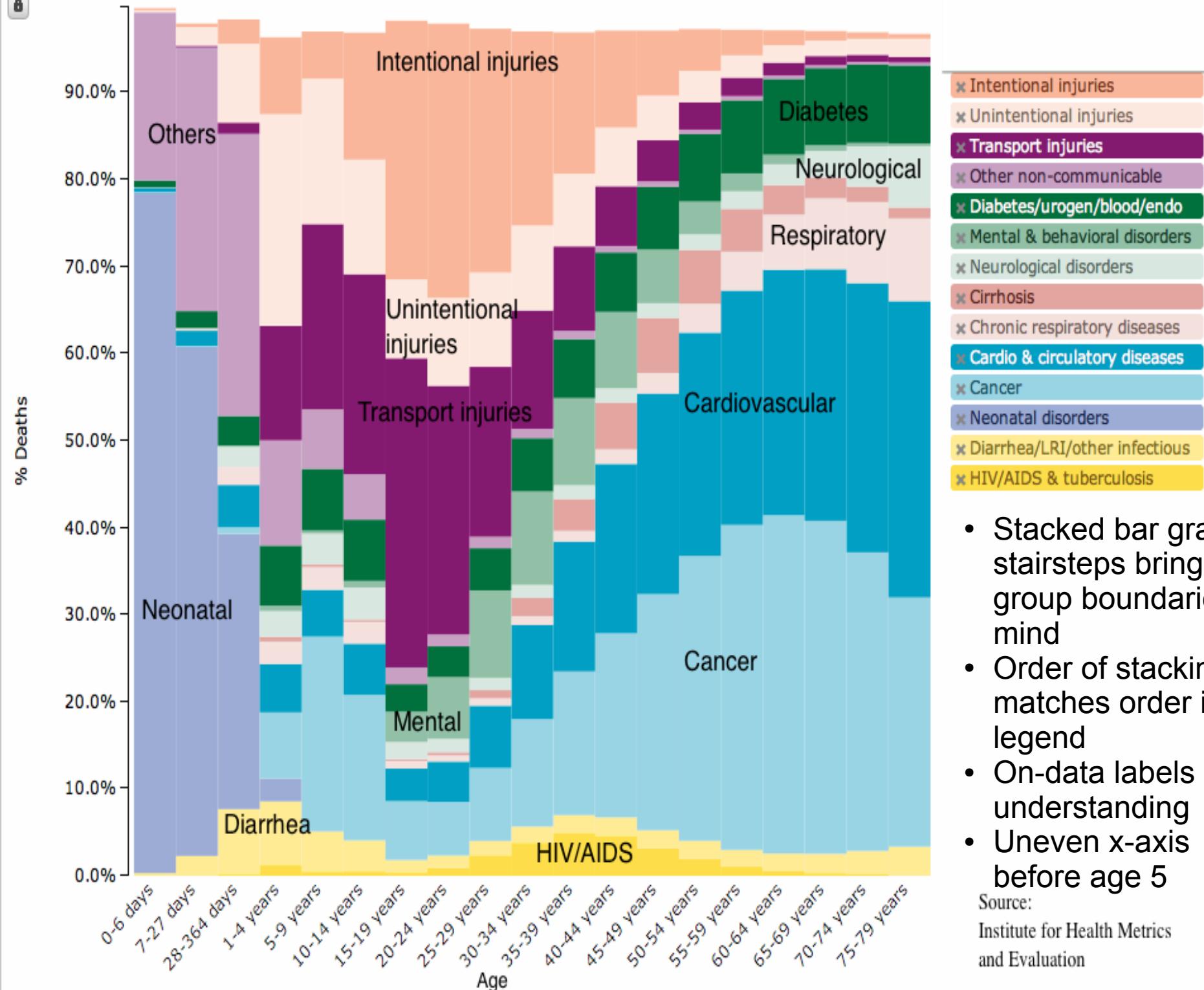


Centers for Disease
Control and Prevention
National Center for Injury
Prevention and Control



- Unlike CDC data, suicide and homicide are combined here under Intentional Injuries.

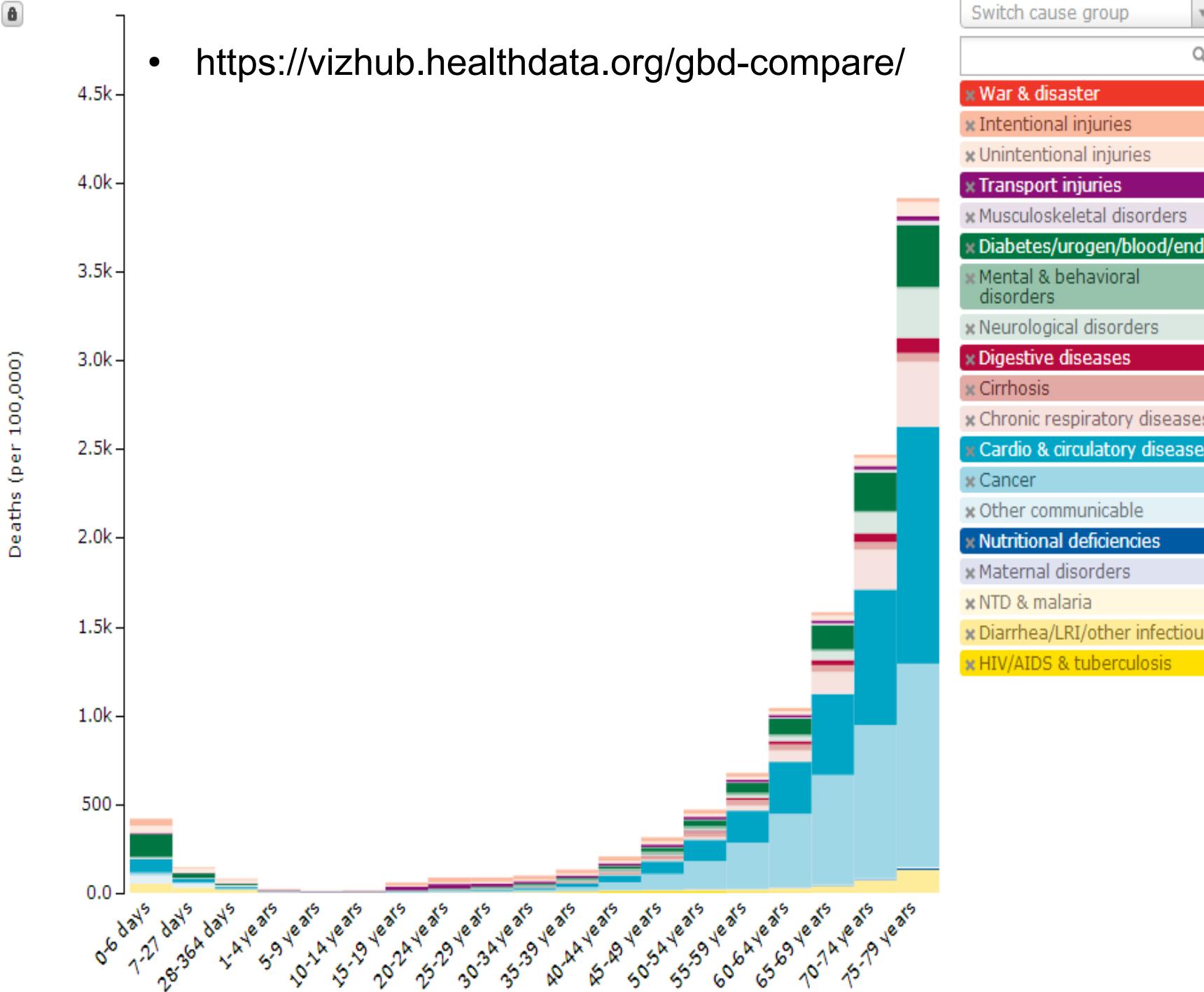
Source:
Institute for Health Metrics
and Evaluation



- Stacked bar graph--stairsteps bring group boundaries to mind
- Order of stacking matches order in legend
- On-data labels ease understanding
- Uneven x-axis before age 5

Source:
Institute for Health Metrics
and Evaluation

- <https://vizhub.healthdata.org/gbd-compare/>

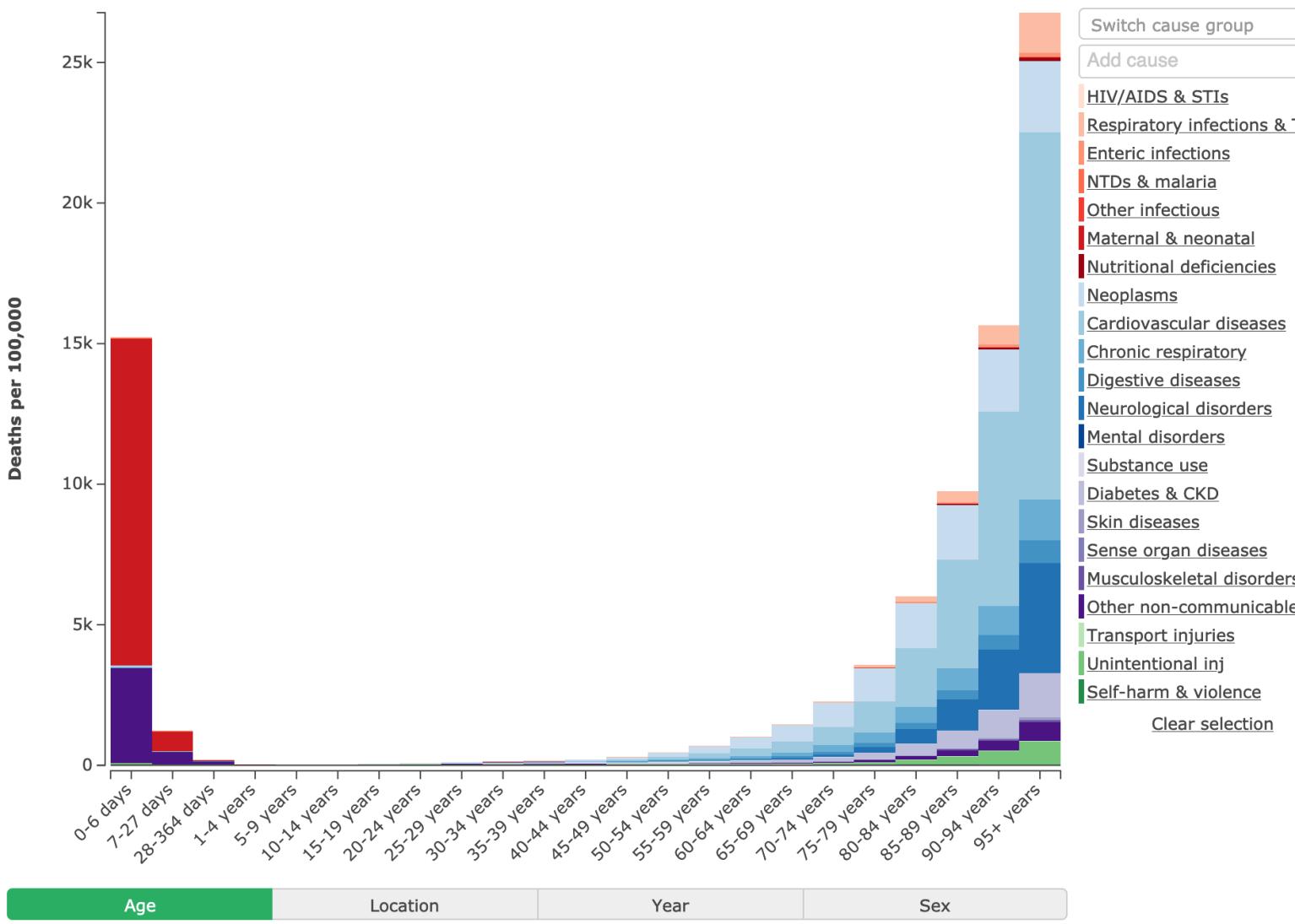


Single

Explore

Compare ▾

High-income North America, Both sexes, 2019

Settings [Use advanced settings](#)**Display** Cause Risk**Measure** Deaths YLDs DALYs**Location** High-income North America**Year** 2019**Sex** Male Female Both**Units** # Rate %[Take tour ►](#)

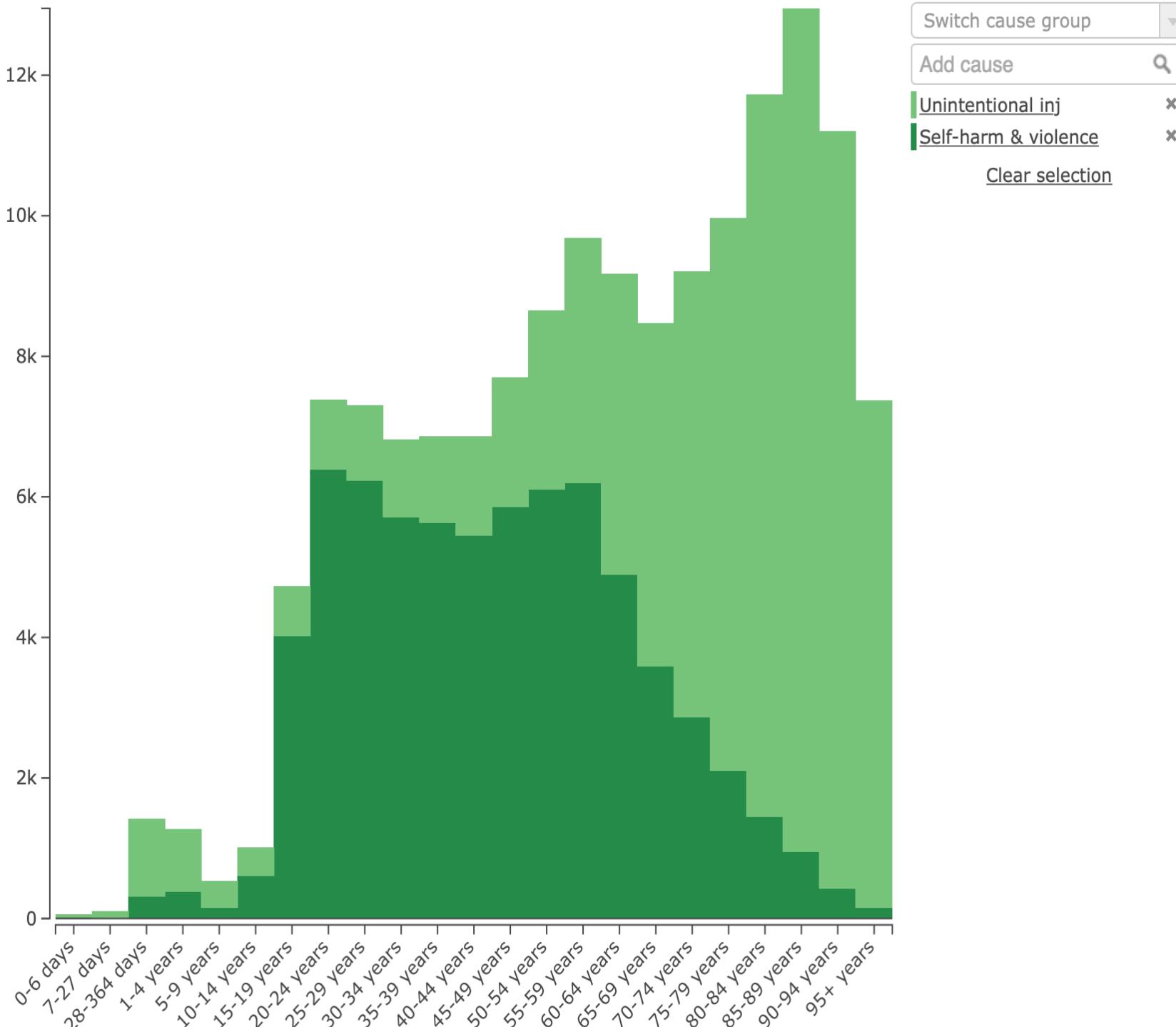
- [Switch cause group](#)
- [Add cause](#)
- HIV/AIDS & STIs
- Respiratory infections & TI
- Enteric infections
- NTDs & malaria
- Other infectious
- Maternal & neonatal
- Nutritional deficiencies
- Neoplasms
- Cardiovascular diseases
- Chronic respiratory
- Digestive diseases
- Neurological disorders
- Mental disorders
- Substance use
- Diabetes & CKD
- Skin diseases
- Sense organ diseases
- Musculoskeletal disorders
- Other non-communicable
- Transport injuries
- Unintentional inj
- Self-harm & violence

[Clear selection](#)

- This graph shows the (browser) control panel
- This graph goes out to age = 95
- Note the colors for the categories --shades of red, blue, purple...

High-income North America, Both sexes, 2019

i



• Units: #

Switch cause group

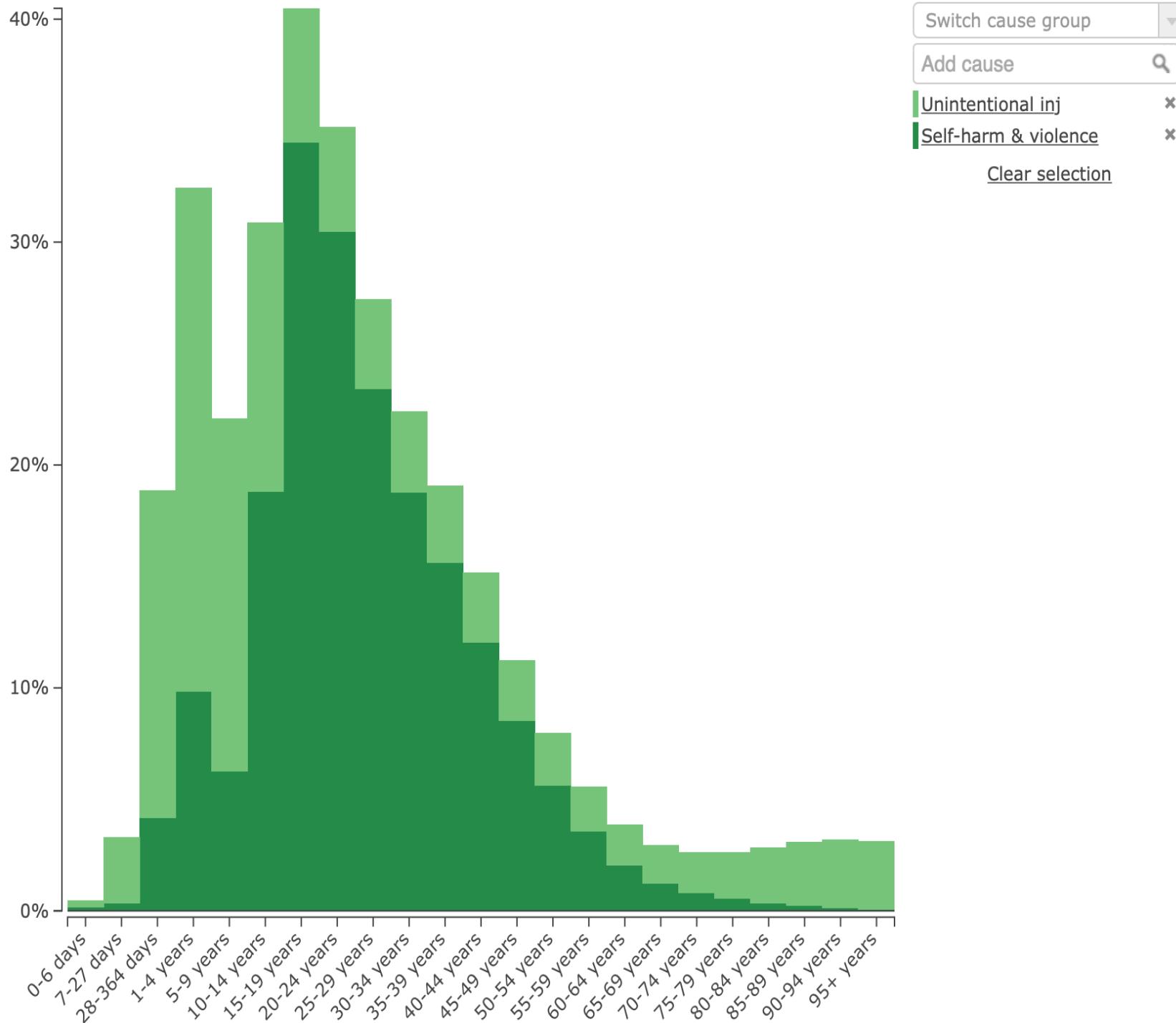
Add cause

Unintentional inj

Self-harm & violence

Clear selection

High-income North America, Both sexes, 2019



• Units: %

Age

Location

Year

Sex

Switch cause group

Add cause

Unintentional inj

Self-harm & violence

Clear selection

High-income North America, Both sexes, 2019



Switch cause group

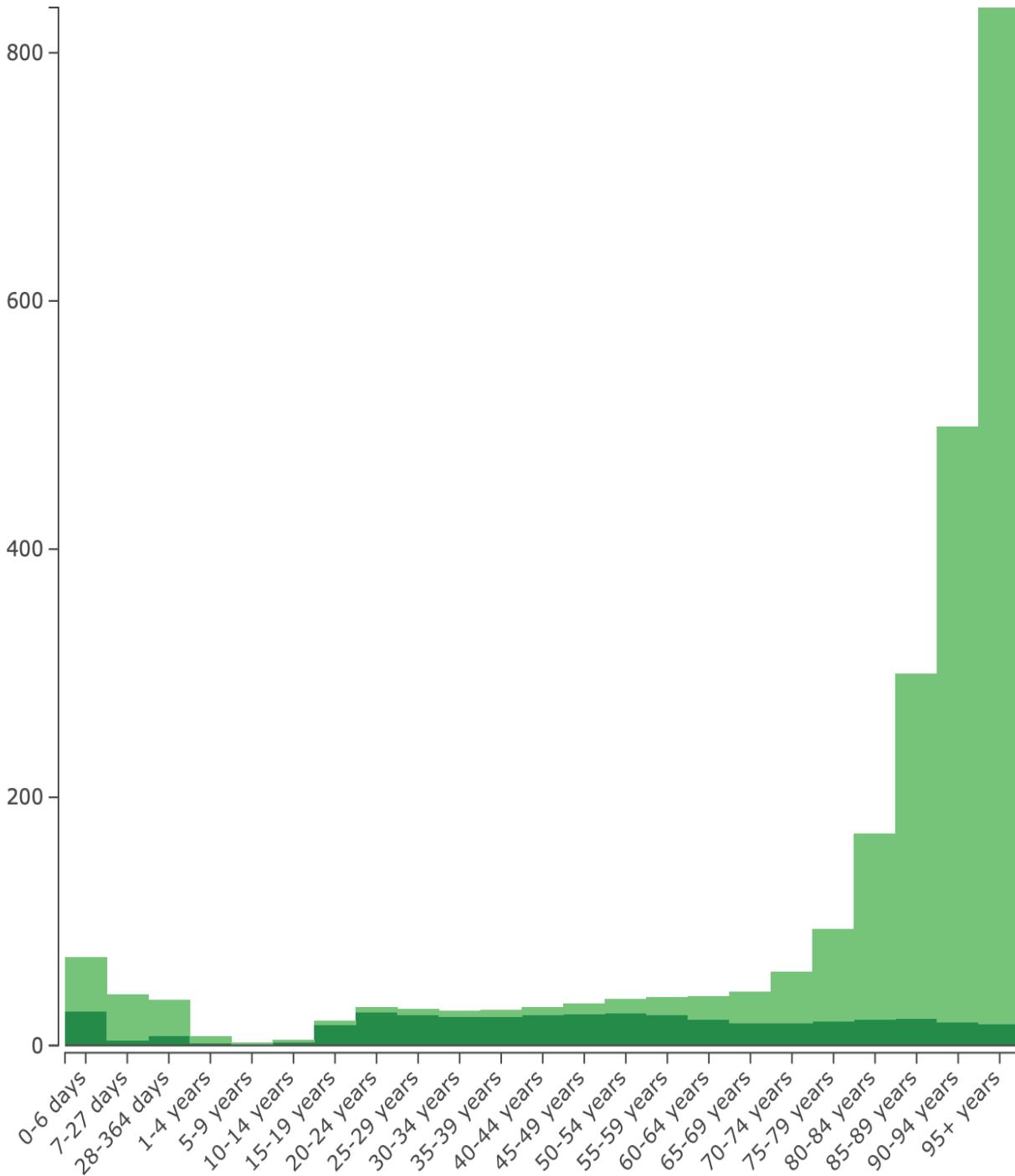
Add cause

Unintentional inj

Self-harm & violence

[Clear selection](#)

Deaths per 100,000



- Units: per 100,000
- We really do need different views of the data to understand it.

Age

Location

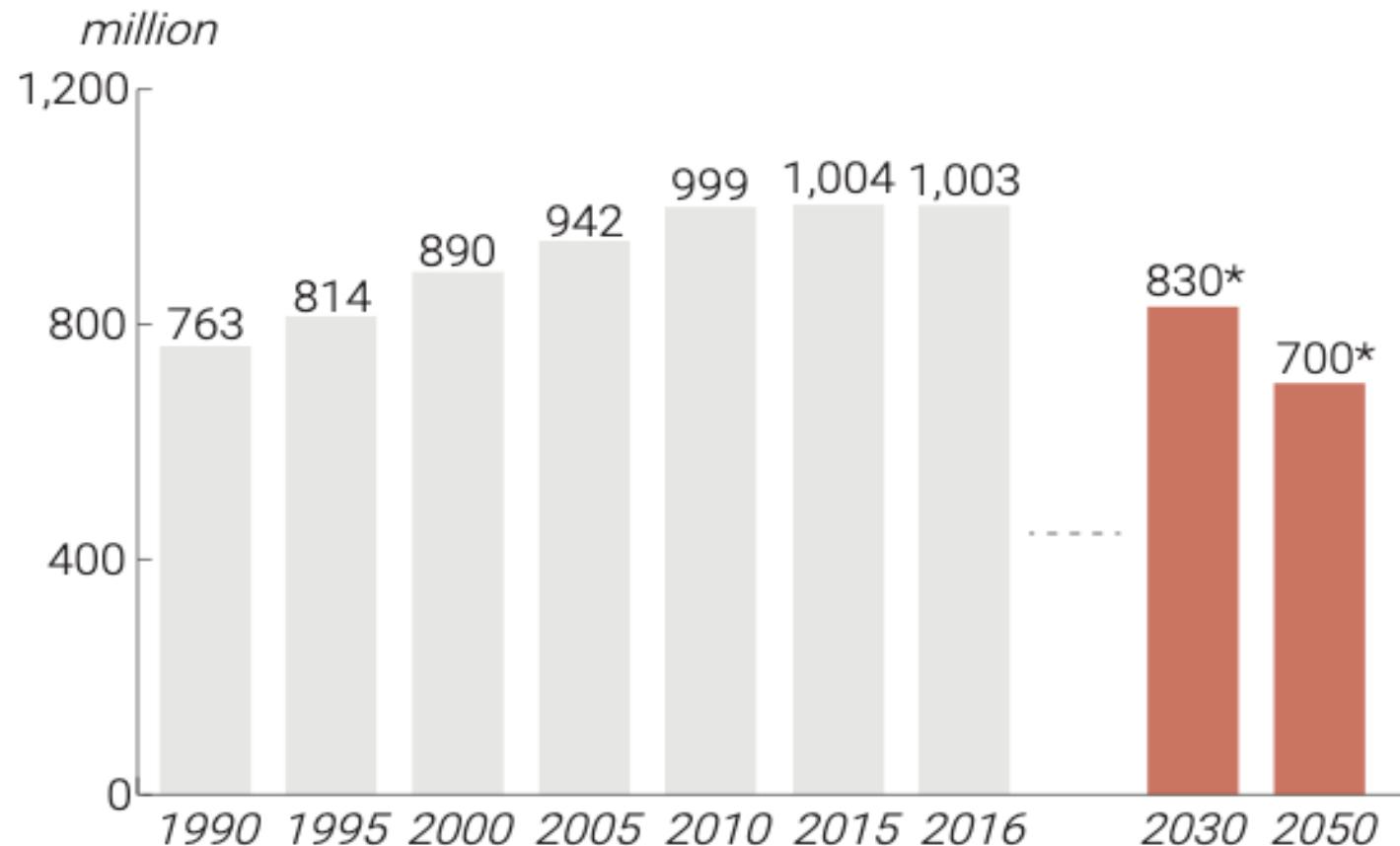
Year

Sex

What's wrong here?

Working age population to decline

15-64 age population of China



*Estimated by Ministry of Human Resources and Social Security

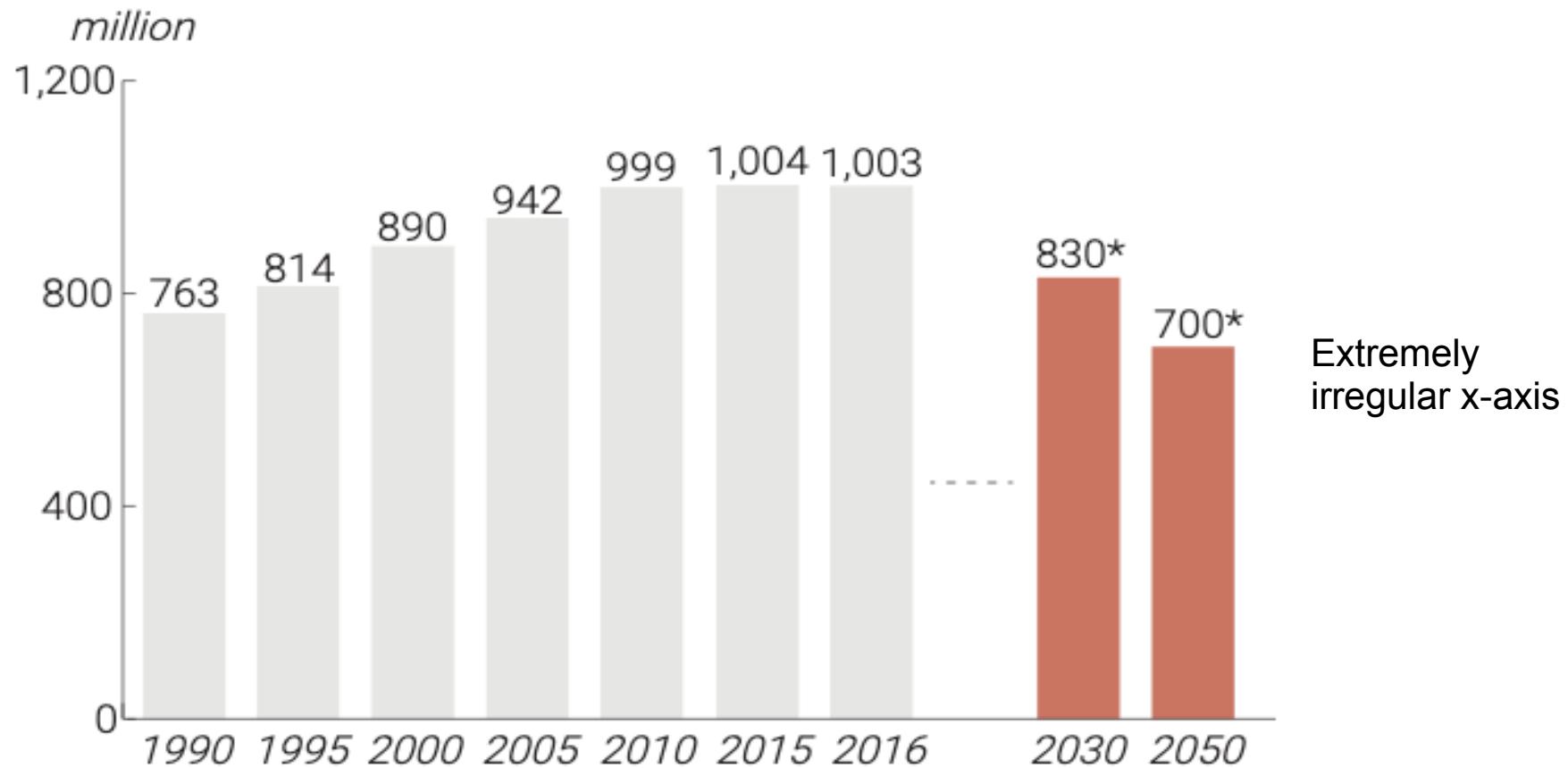
Why 'Made in China 2025' triggered the wrath of President Trump,

Alice Tse and Julianna Wu South China Morning Post 2018-09-11

What's wrong here?

Working age population to decline

15-64 age population of China



*Estimated by Ministry of Human Resources and Social Security

Why 'Made in China 2025' triggered the wrath of President Trump,

Alice Tse and Julianna Wu South China Morning Post 2018-09-11

Activity: Representations

In one minute,
how many ways you can write or draw

75 and 37

Activity: Representations

Written (Arabic, Roman, tally marks)

Length

Area

Arc, fraction

Counting objects

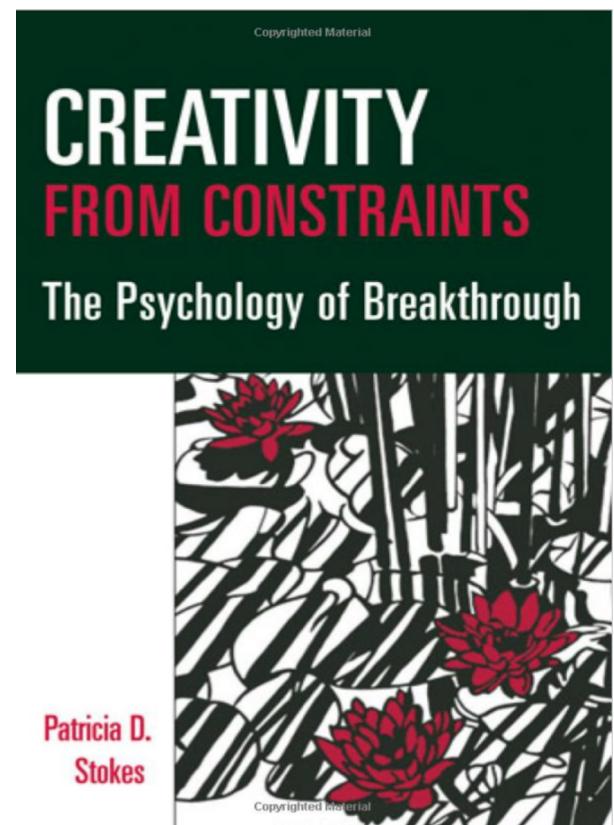
Activity: Representations

How can we generate more?

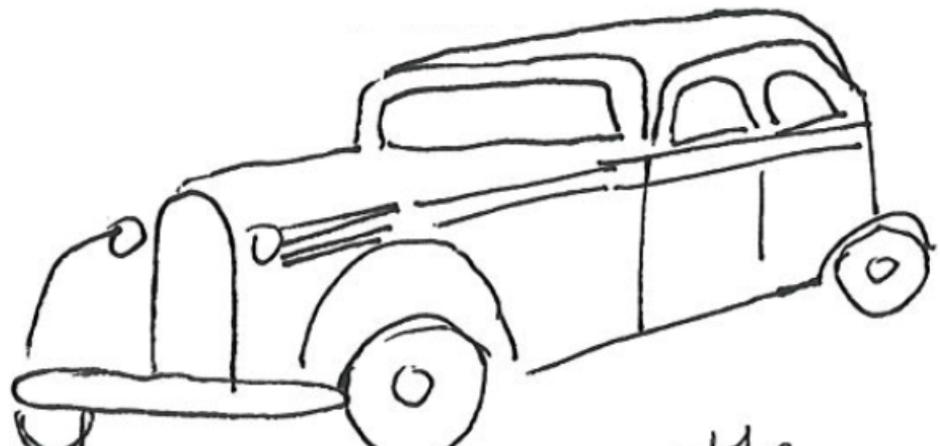
What strategies do we have to help
generate new ideas?

Activity: Representations

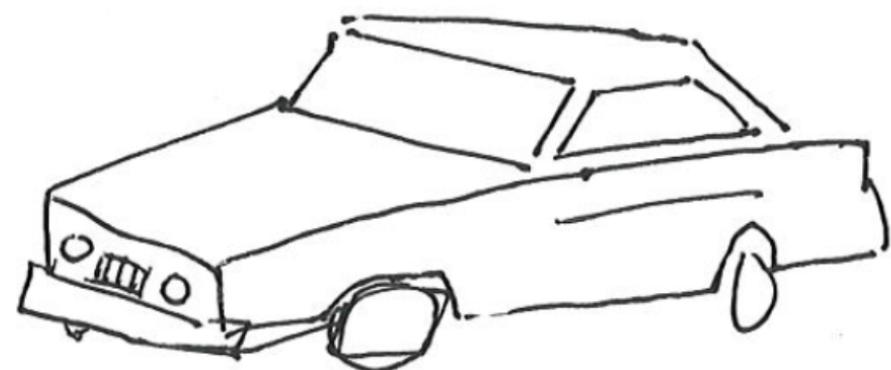
- Early exposure to examples
- Rapid generation of low-value ideas (brainstorming)
- Apply constraint (e.g. only use black-and-white; only use curved lines; only use vertical lines)



Time...



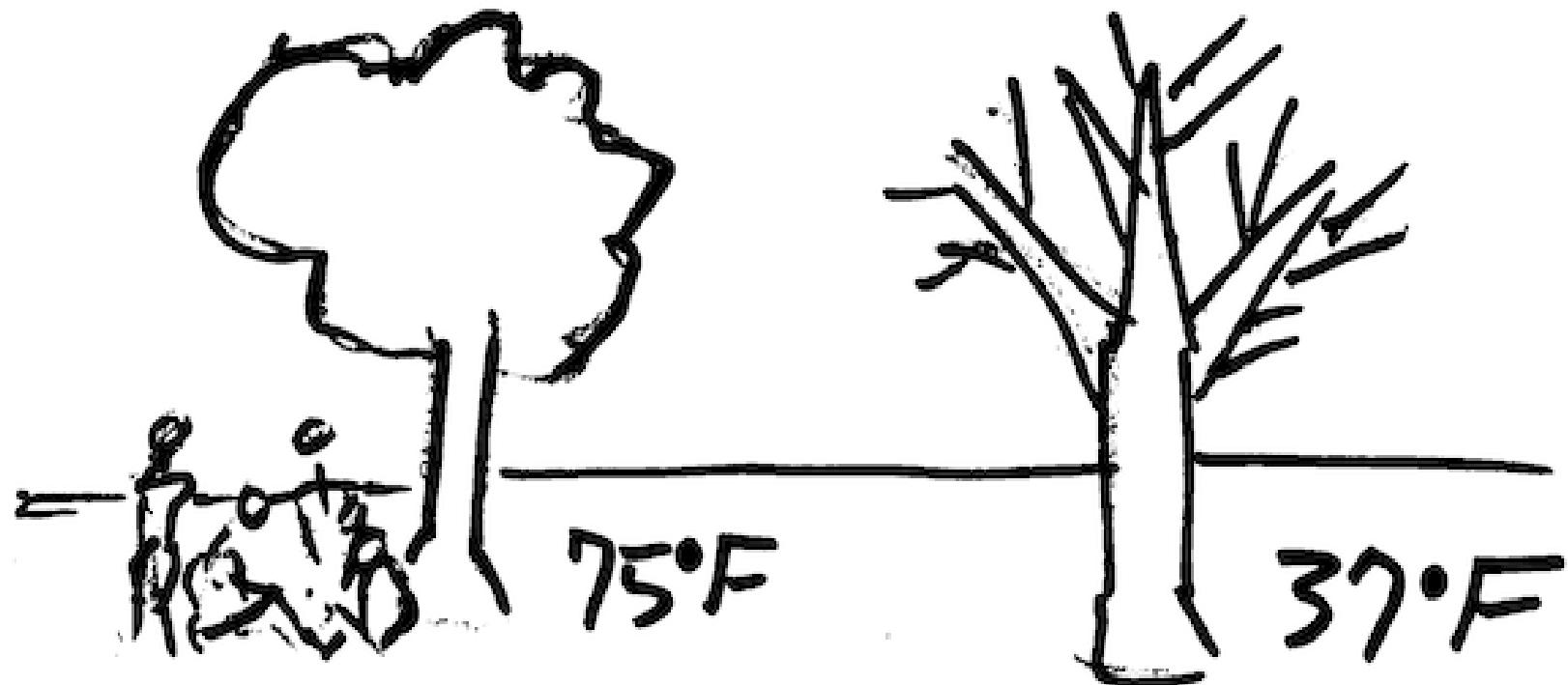
1937 Plymouth



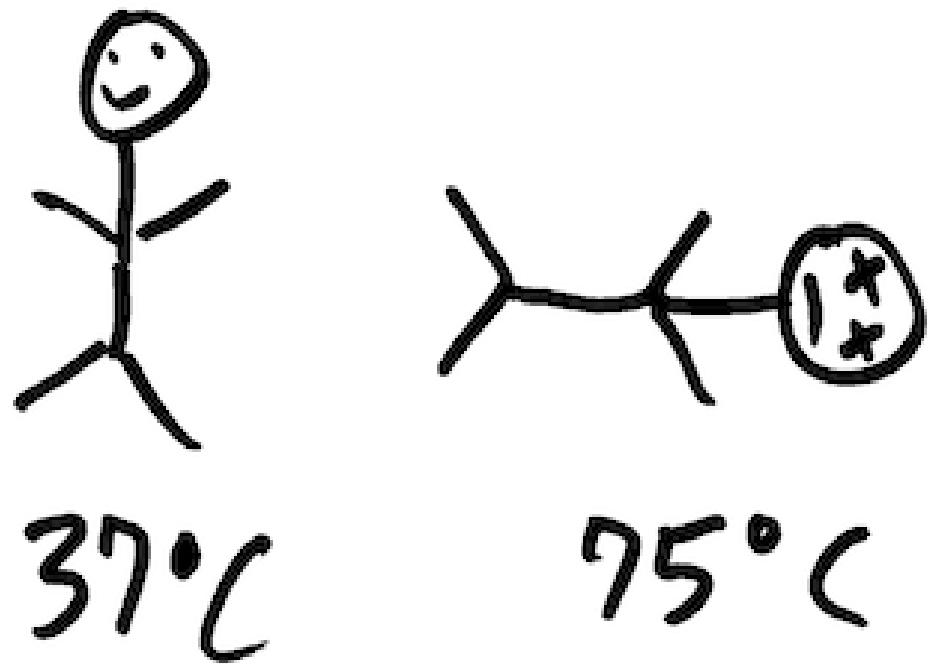
1975 Plymouth

Image credit: Jon Schwabish.

Temperature



Temperature (civilized)

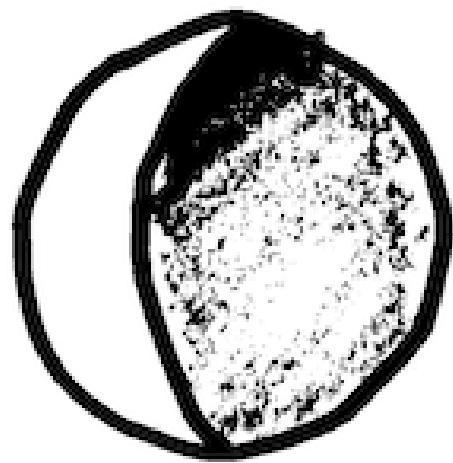


Arrow of time

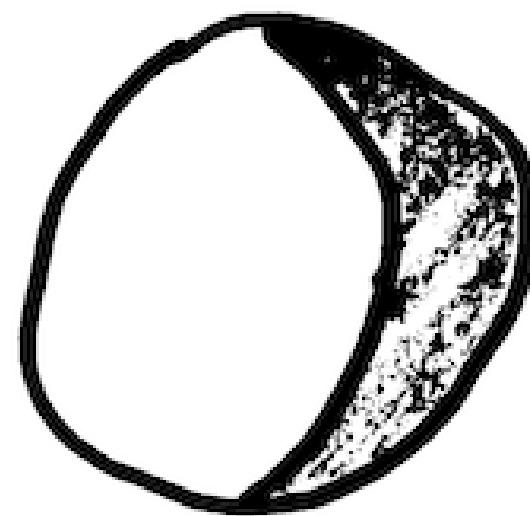


37

75

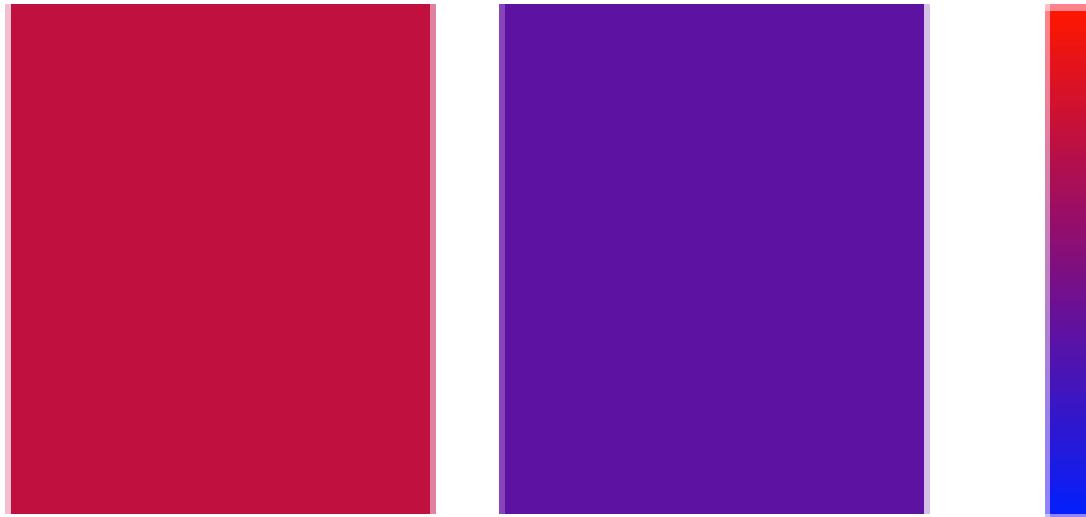


37



75

Color?



<https://rockcontent.com/blog/45-ways-to-communicate-two-quantities/>

Santiago Ortiz

<https://rockcontent.com/blog/45-ways-to-communicate-two-quantities/>

Encoding

Data

- physical data types
(int, float, double, string)
- conceptual data types
(date, temperature, price)

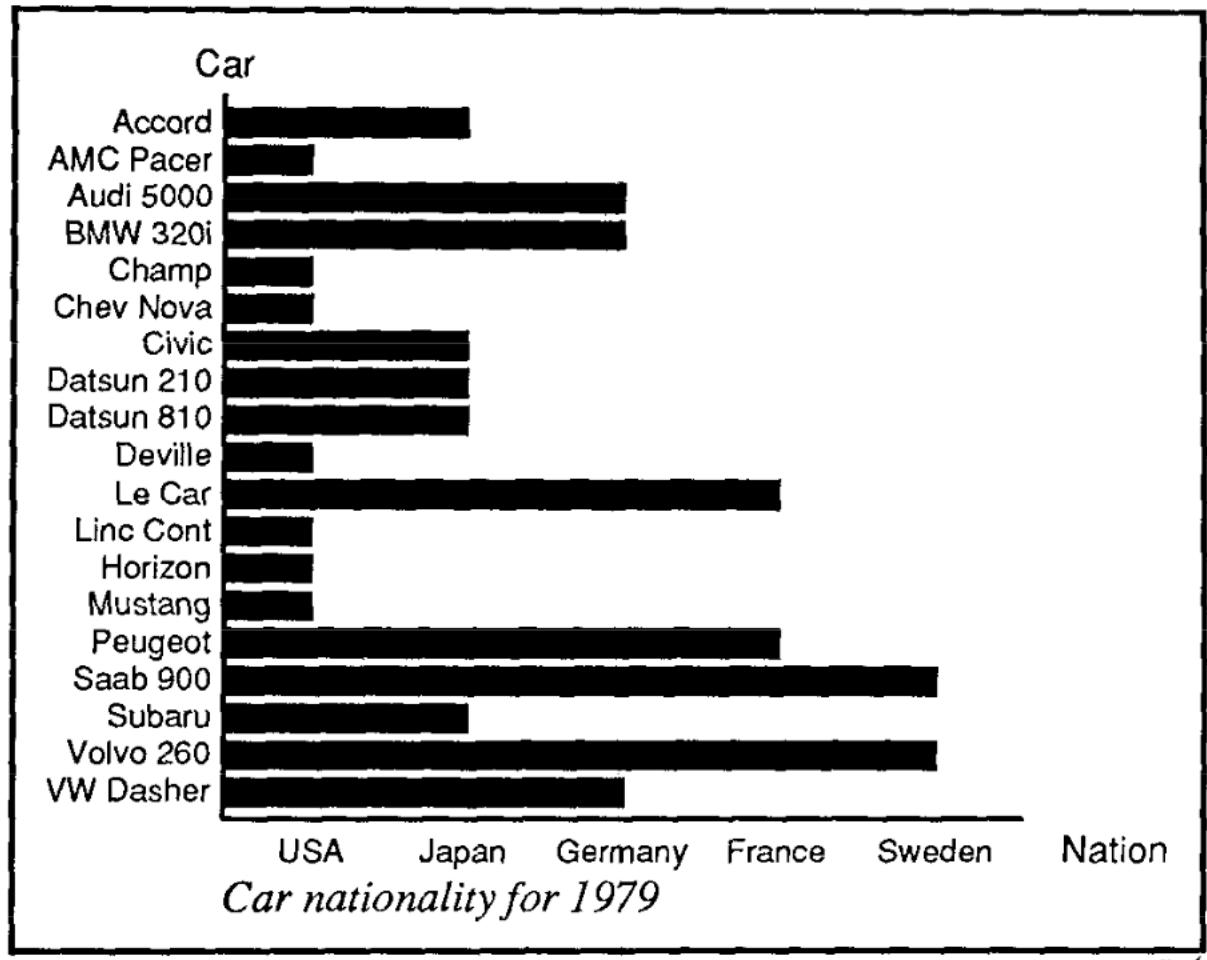
Graphic

- visual channels
(position, color, opacity)
- visual marks
(line, point, rectangle, area)

Principles

“The **expressiveness principle** dictates that the visual encoding should express **all of, and only, the information** in the dataset attributes.”

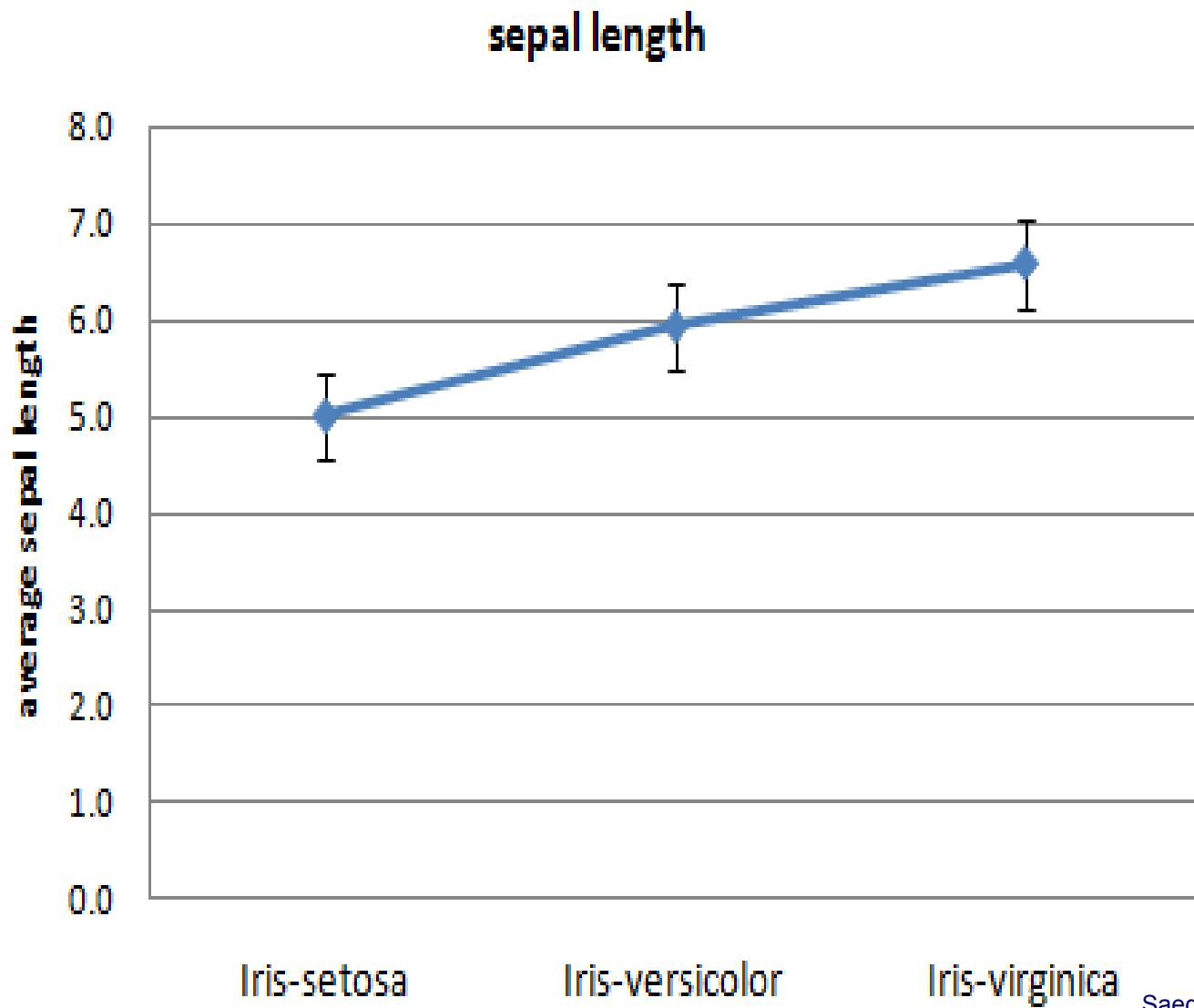
What's wrong here?



apt •

Mackinlay, 1986, Automating the Design of Graphical Presentations of Relational Information

Don't do this!

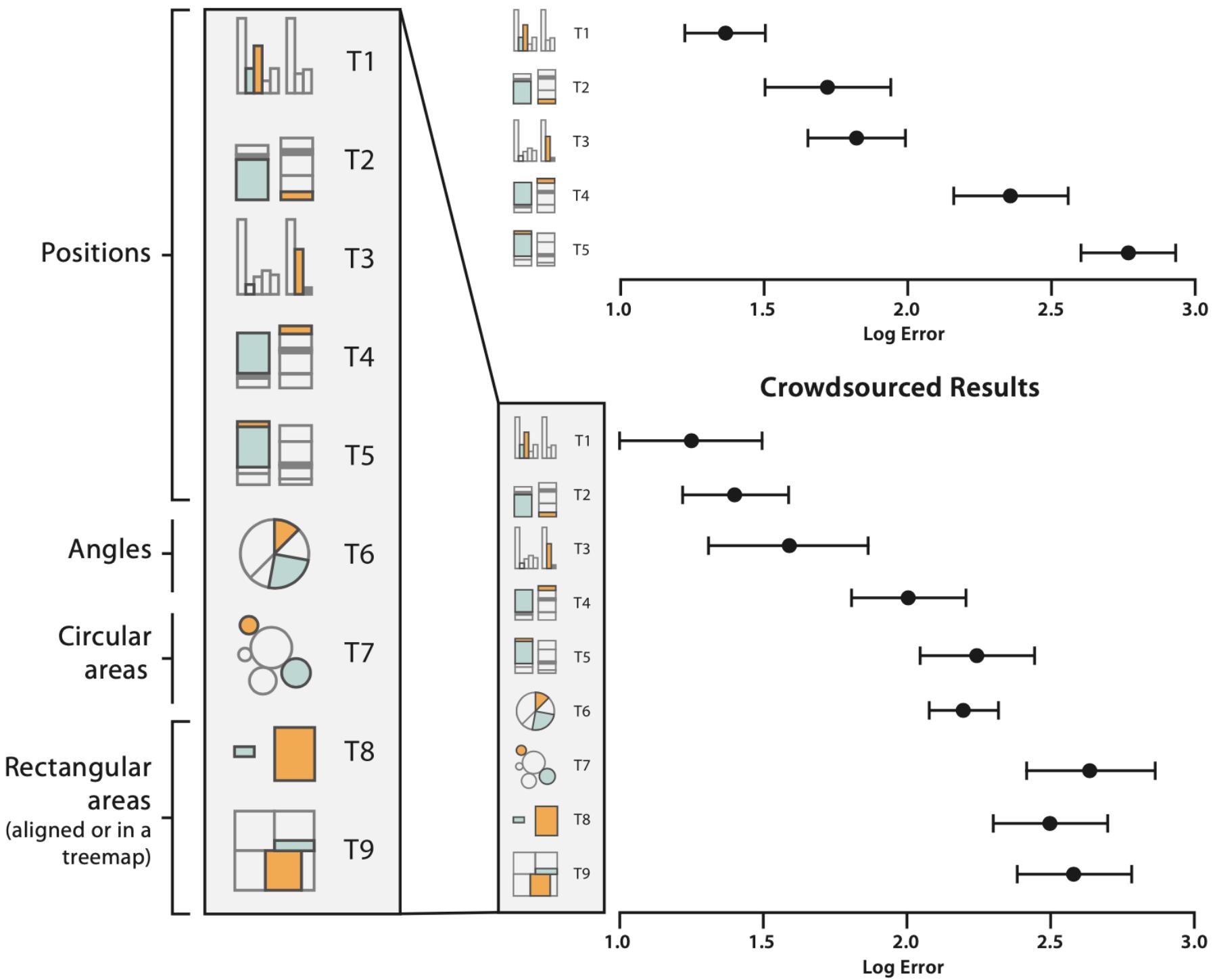


Principles

“The **expressiveness principle** dictates that the visual encoding should express all of, and only, the information in the dataset attributes.”

“The **effectiveness principle** dictates that the importance of the attribute should match the **salience of the channel**”

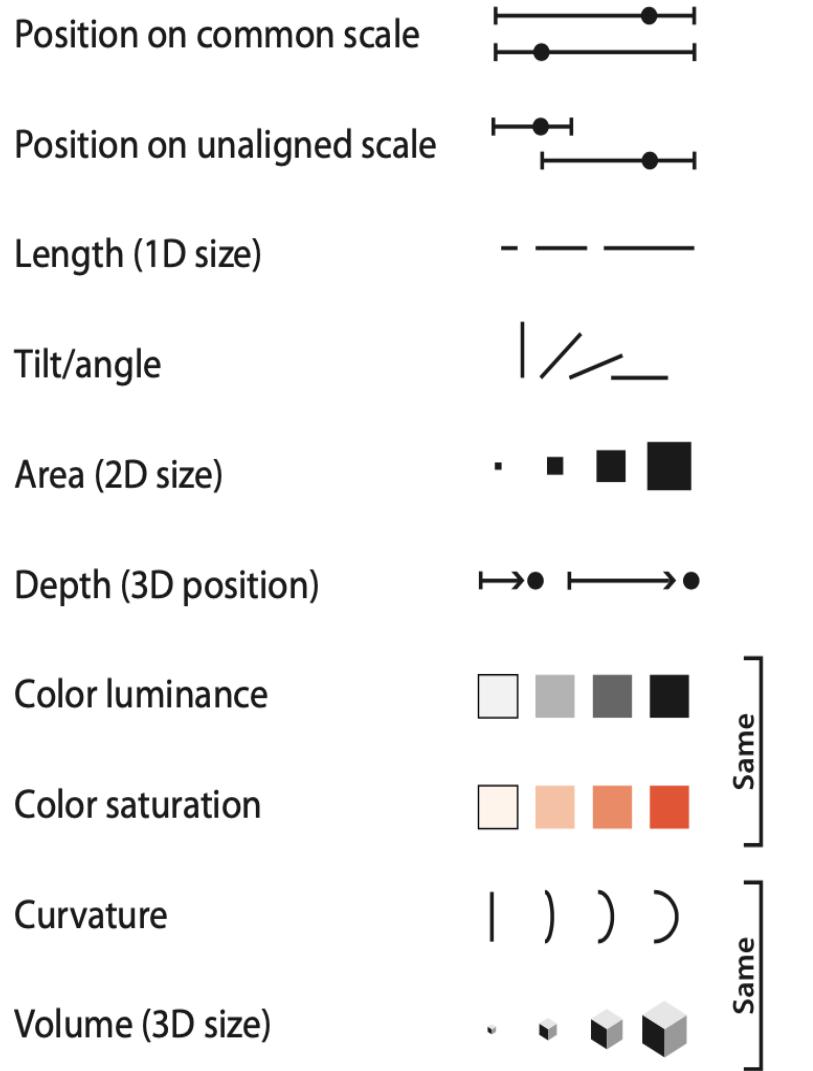
Cleveland & McGill's Results



Encoding channels

Channels: Expressiveness Types and Effectiveness Ranks

- **Magnitude Channels:** O and Q



- **Identity Channels :** N



Graphical channel real estate is precious.

- Munzner, Visualization Analysis and Design, 2014

Encoding

A **mark** is a basic graphical element in an image.

A visual channel is a way to control the appearance of marks,

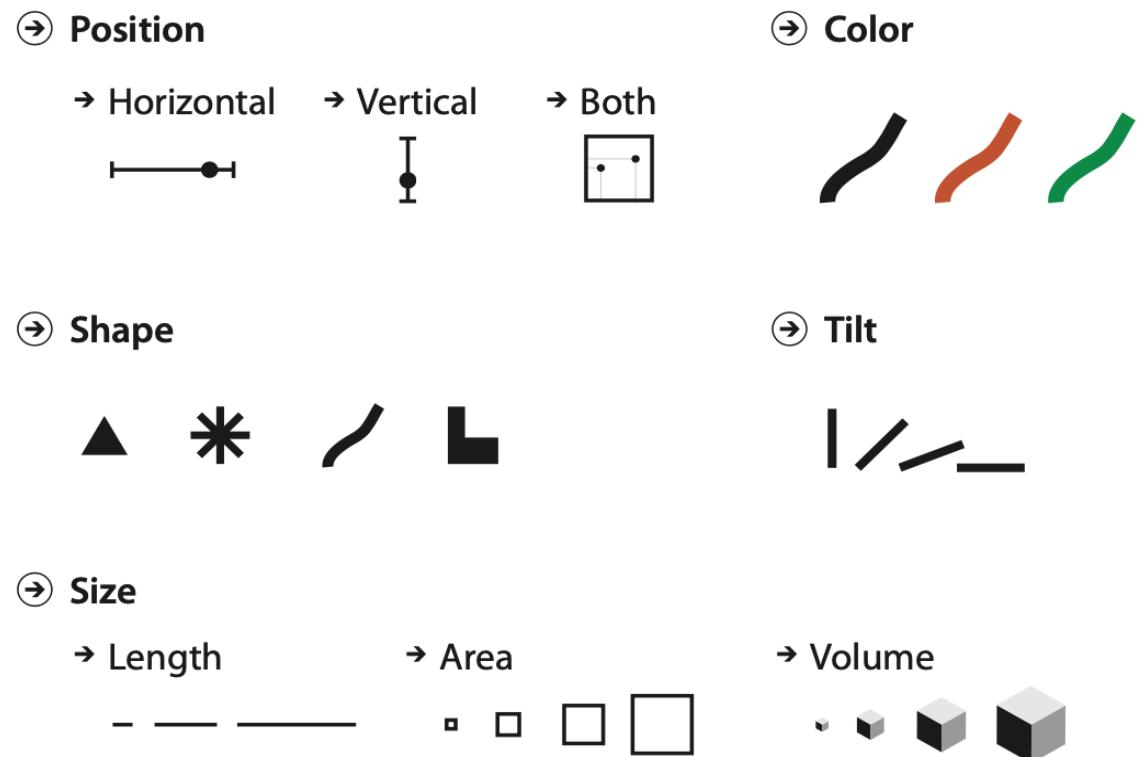
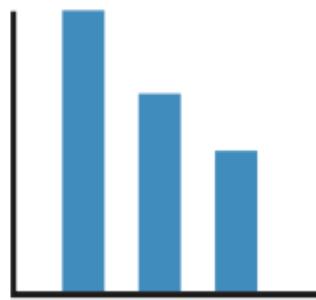
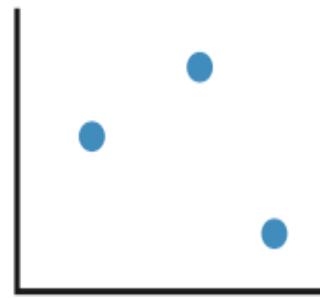


Figure 5.3. Visual channels control the appearance of marks.

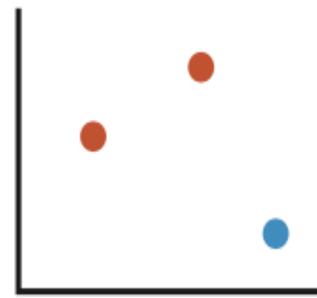
How many channels?



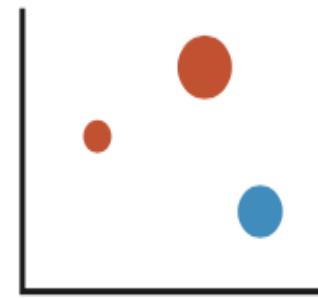
(a)



(b)

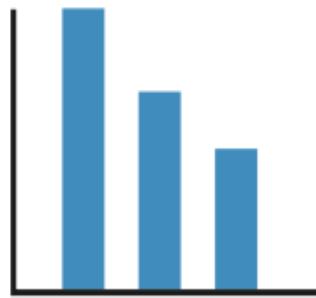


(c)

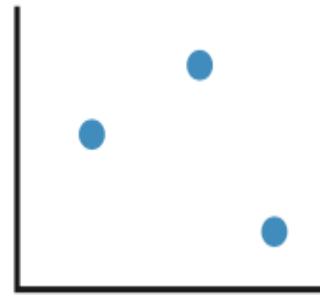


(d)

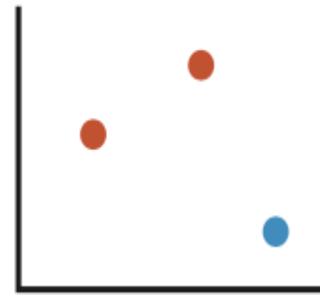
How many channels?



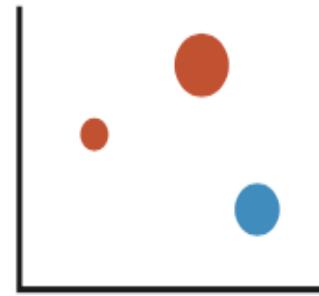
(a)



(b)



(c)



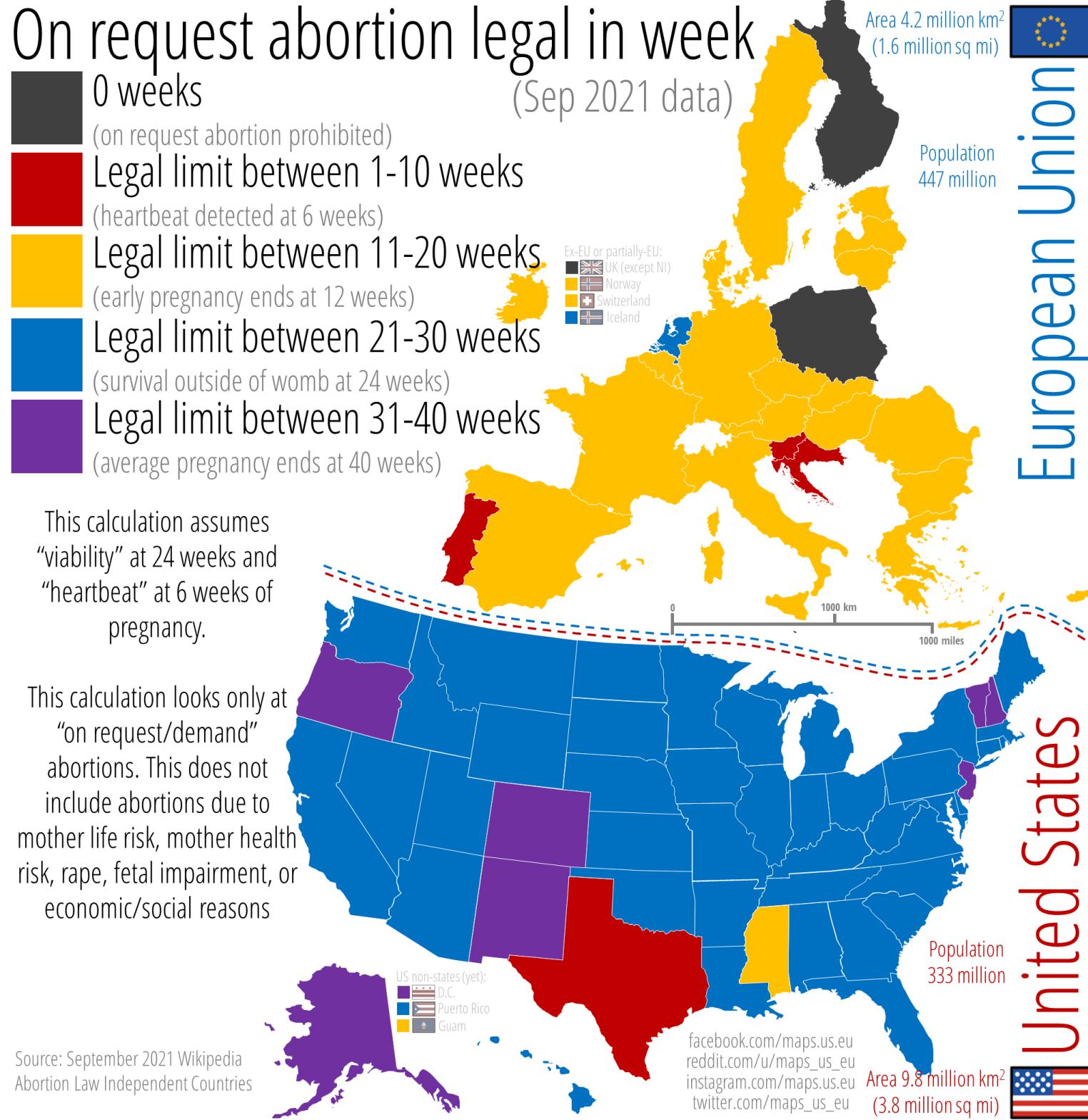
(d)

2

2

3

4



On request abortion legal in week

0 weeks

(on request abortion prohibited)

Legal limit between 1-10 weeks

(heartbeat detected at 6 weeks)

Legal limit between 11-20 weeks

(early pregnancy ends at 12 weeks)

Legal limit between 21-30 weeks

(survival outside of womb at 24 weeks)

Legal limit between 31-40 weeks

(average pregnancy ends at 40 weeks)

This calculation assumes
“viability” at 24 weeks and
“heartbeat” at 6 weeks of
pregnancy.

This calculation looks only at
“on request/demand”
abortions. This does not
include abortions due to
mother life risk, mother health
risk, rape, fetal impairment, or
economic/social reasons

Area 4.2 million km²
(1.6 million sq mi)



European Union

Population
447 million

Ex-EU or partially-EU:
UK (except NI)
Norway
Switzerland
Iceland

0 1000 km

Where is Norway?
And the UK???

United States



Population
333 million

Area 9.8 million km²
(3.8 million sq mi)

facebook.com/maps.us.eu
reddit.com/u/maps_us_eu
instagram.com/maps.us.eu
twitter.com/maps_us_eu

US non-states (yet):
D.C.
Puerto Rico
Guam

Source: September 2021 Wikipedia
Abortion Law Independent Countries

On request abortion legal in week

Area 4.2 million km²
(1.6 million sq mi)



0 weeks

(on request abortion prohibited)

Legal limit between 1-10 weeks

(heartbeat detected at 6 weeks)

Legal limit between 11-20 weeks

(early pregnancy ends at 12 weeks)

Legal limit between 21-30 weeks

(survival outside of womb at 24 weeks)

Legal limit between 31-40 weeks

(average pregnancy ends at 40 weeks)

This calculation assumes
“viability” at 24 weeks and
“heartbeat” at 6 weeks of
pregnancy.

This calculation looks only at
“on request/demand”
abortions. This does not
include abortions due to
mother life risk, mother health
risk, rape, fetal impairment, or
economic/social reasons

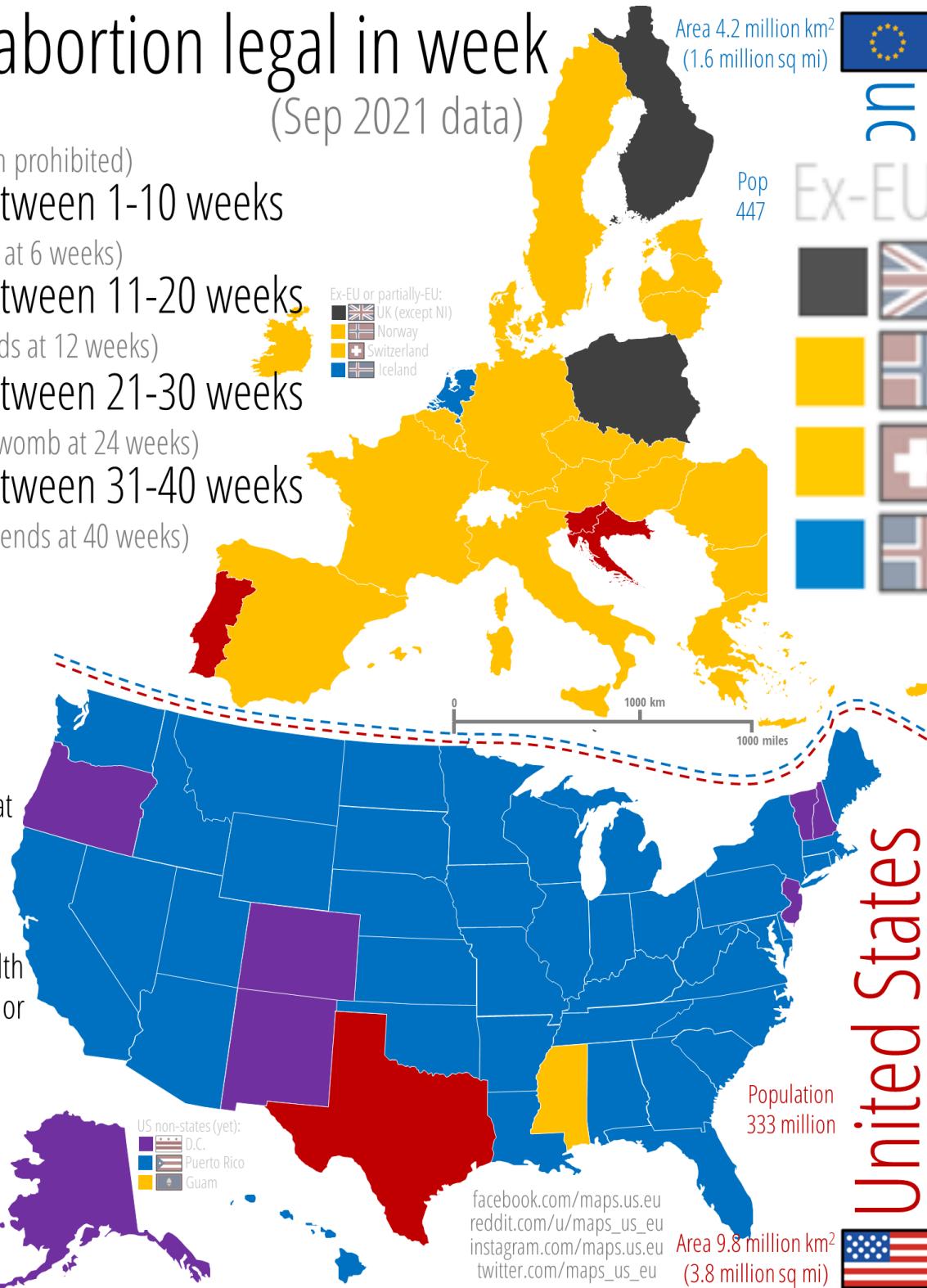
(Sep 2021 data)

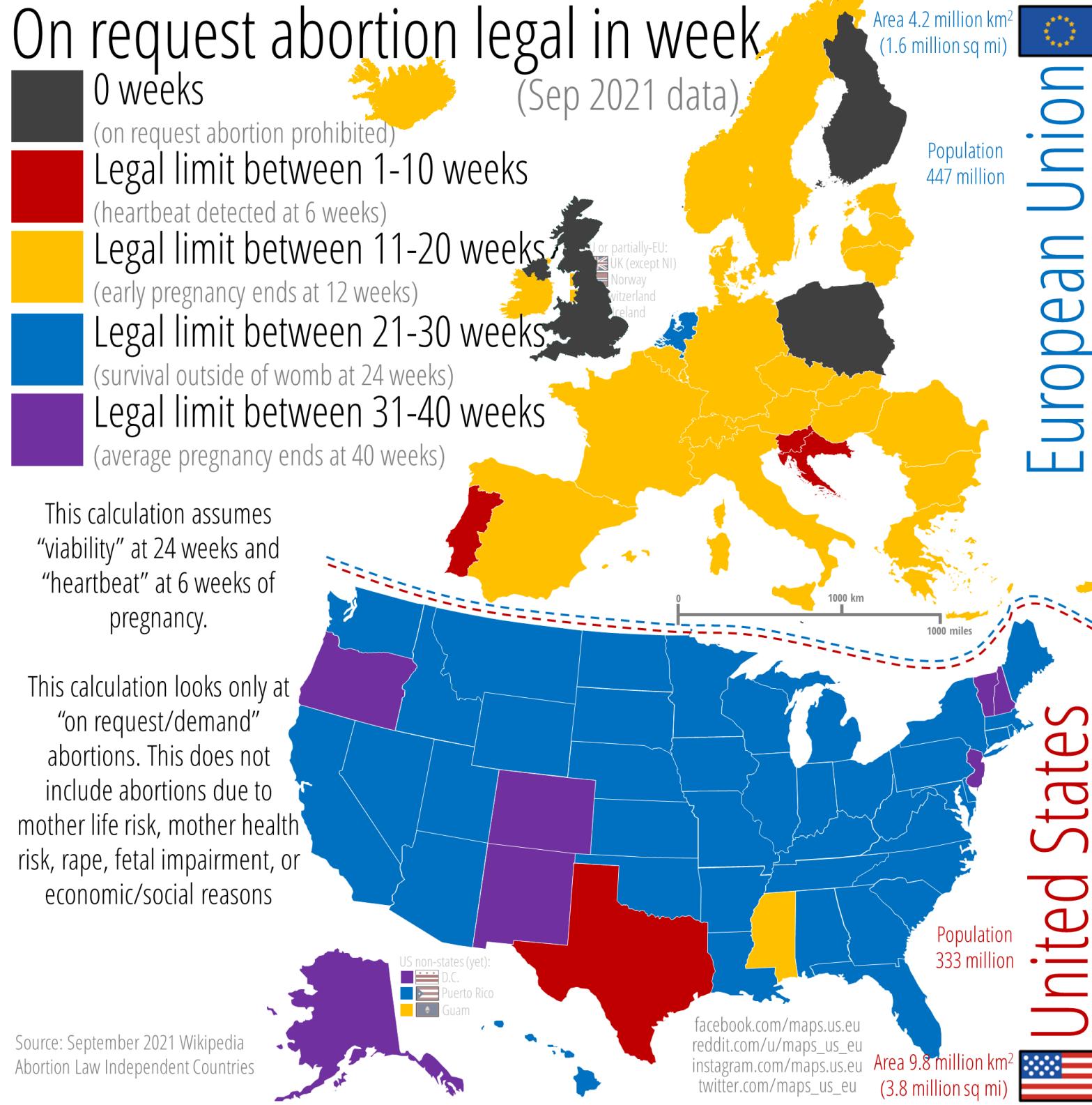
Pop
447

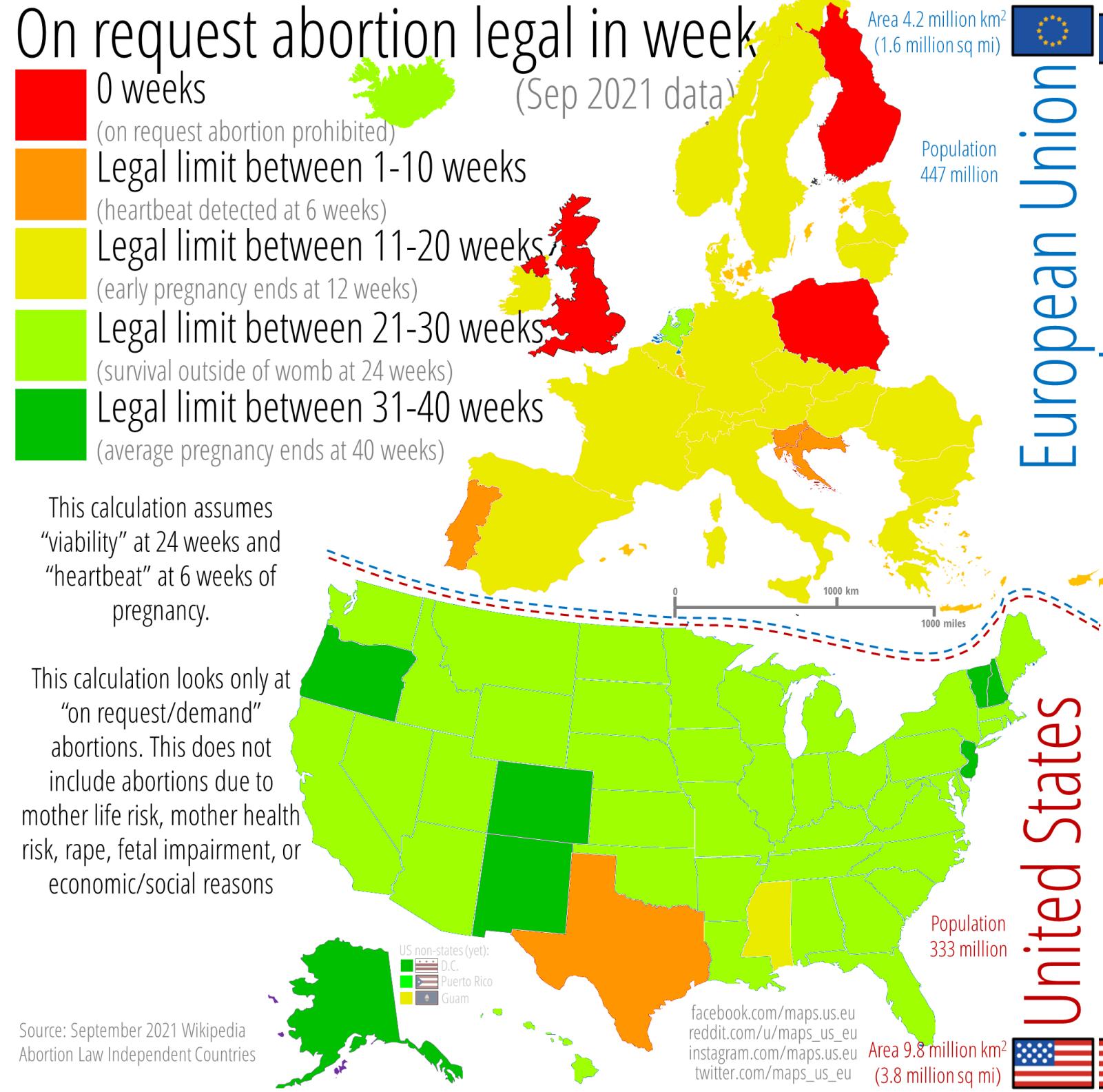
Ex-EU or partially-EU:
UK (except NI)
Norway
Switzerland
Iceland

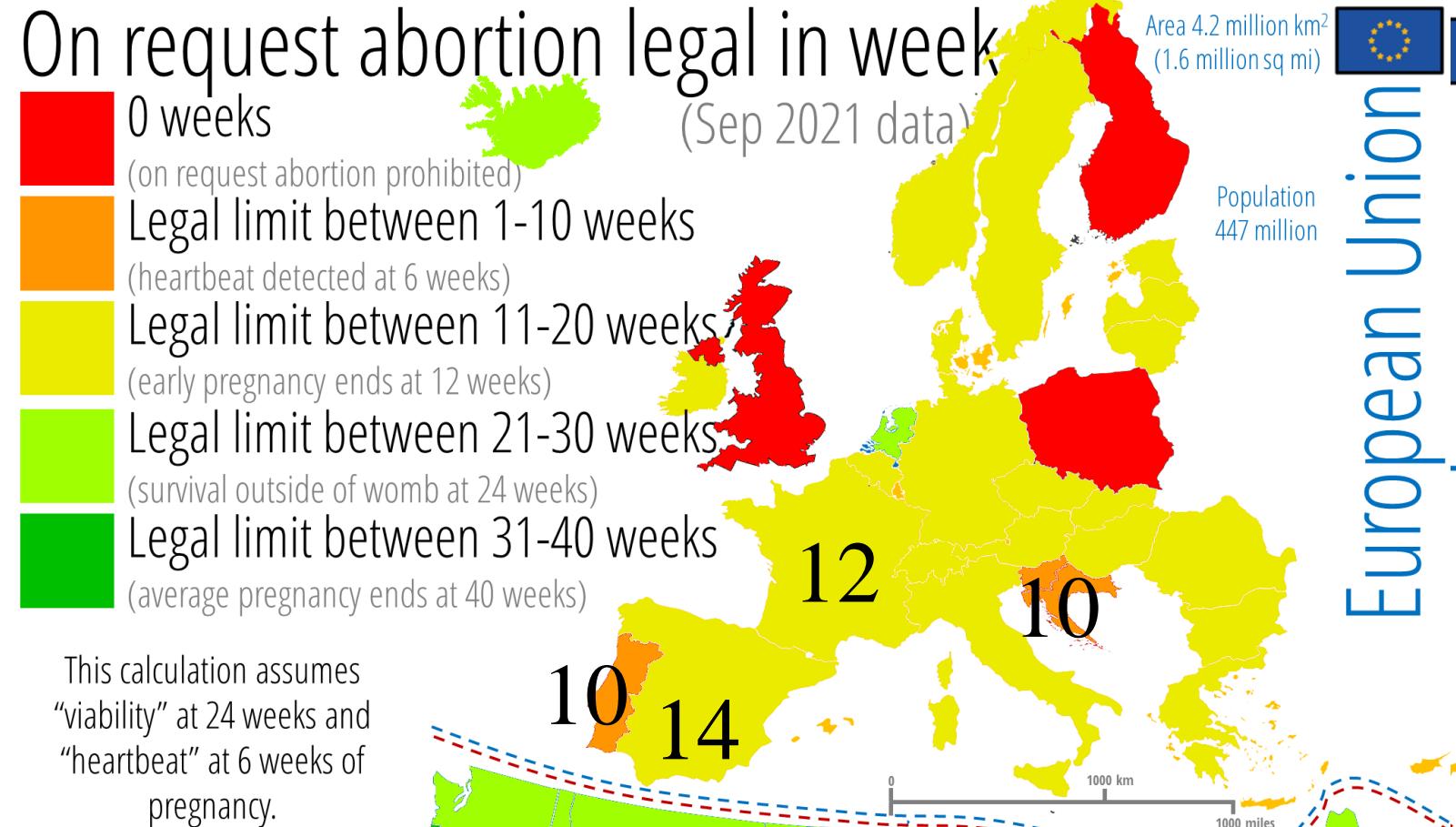


Ex-EU or partially-EU:
UK (except NI)
Norway
Switzerland
Iceland









You can book an appointment with a gynecologist in PT, but not in TX.

Why 1-10 week boundaries?



What did we see?

- There were choices of how to present geography
- Former EU / Schengen Area countries had data, but were in greyed-out text, not country outlines! (Don't do this to your audience, please)
- Scale was ordinal, but color mapping black-red-yellow-blue-purple was hard to understand
- Boundaries at 0, 1-10, 11-20, 21-30 are not without consequences: 6 is more like 0 than 10.
- Maaybe making Texas stand out in red was purposeful.

 Find a State or Territory

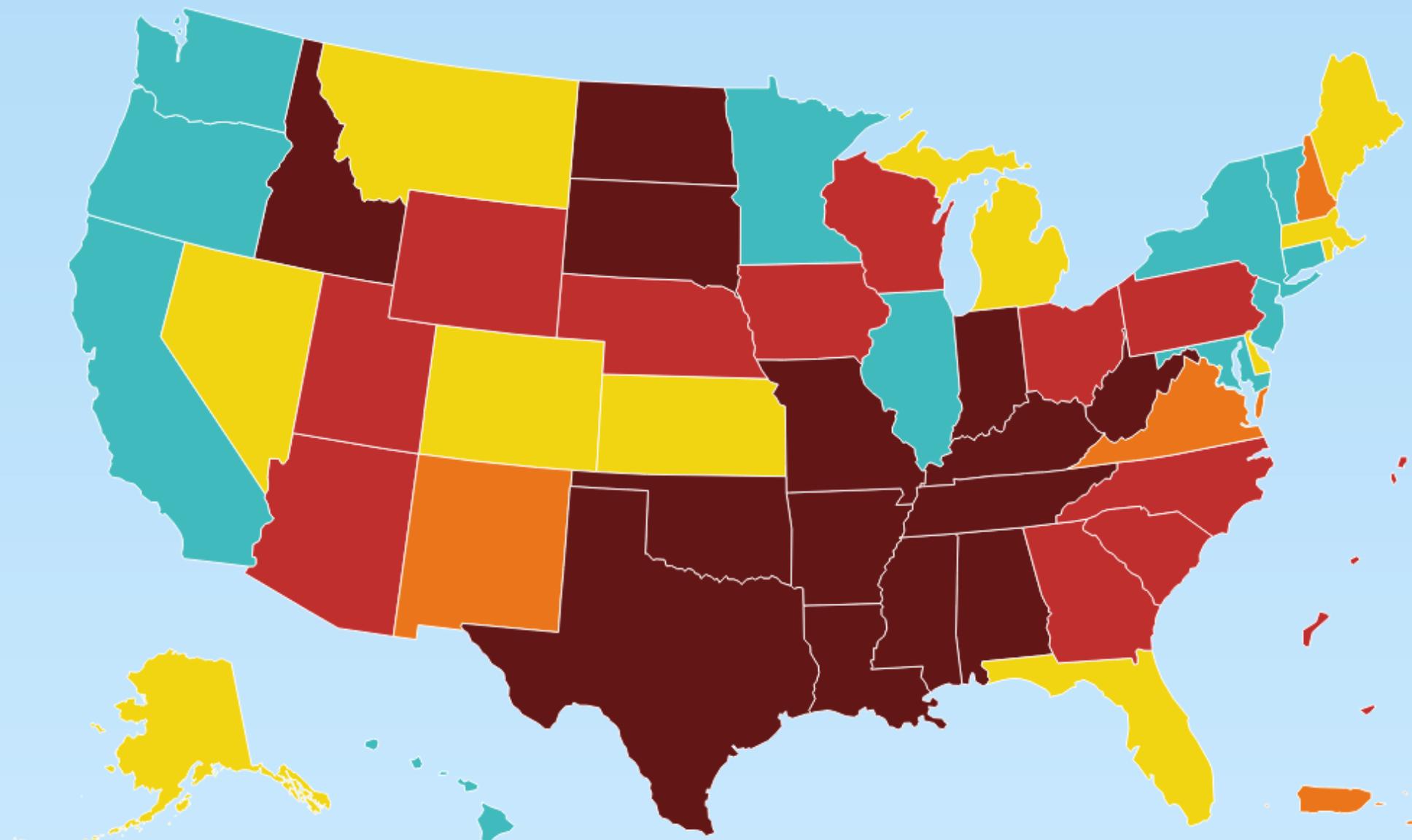
Expanded Access

Protected

Not Protected

Hostile

Illegal



no byline, "After Roe Fell: Abortion Laws by State" Center for Reproductive Rights (current September 2023)
<https://reproductiverights.org/maps/abortion-laws-by-state/>

...and on to hacking

- Notebook on importing census data into pandas, making rudimentary bar graphs with matplotlib.
- `altair_introduction.ipynb`
- Population estimates with single-birth-year resolution:
<https://www.census.gov/data/tables/time-series/demo/popest/pre-1980-national.html>
- <https://www.census.gov/dataviz/visualizations/055/>

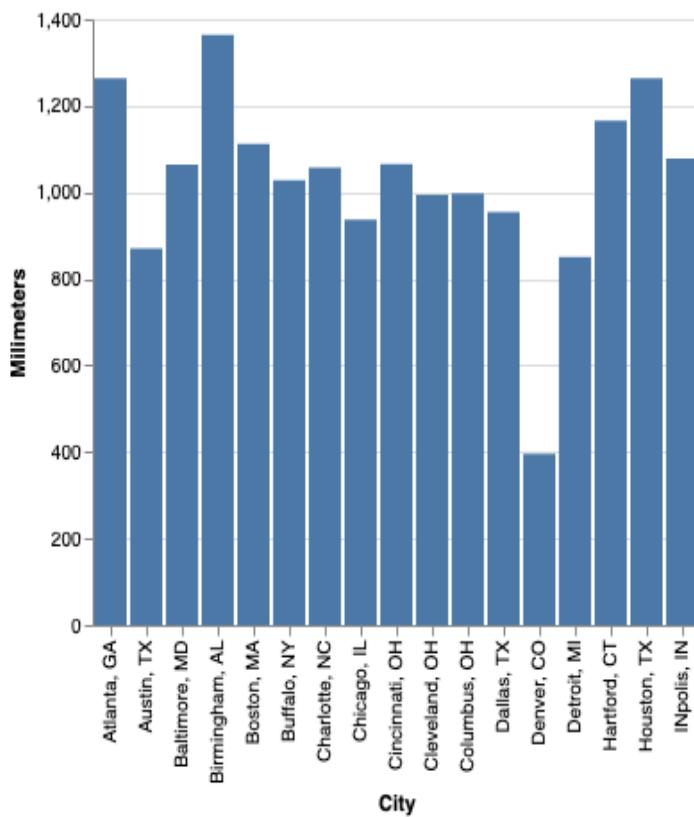
Toy categorical data

- Days City Inches Mm
- 113 Atlanta, Georgia 49.7 1263
- 88 Austin, Texas 34.2 870
- 116 Baltimore, Maryland 41.9 1064
- 117 Birmingham, Alabama 53.7 1364
- 126 Boston, Massachusetts 43.8 1112
- 167 Buffalo, New York 40.5 1028
- 110 Charlotte, North Carolina 41.6 1057
- 124 Chicago, Illinois 36.9 937

Toy data, bar charts

Nominal (city name) => x
Quantitative (Annual precipitation) => y

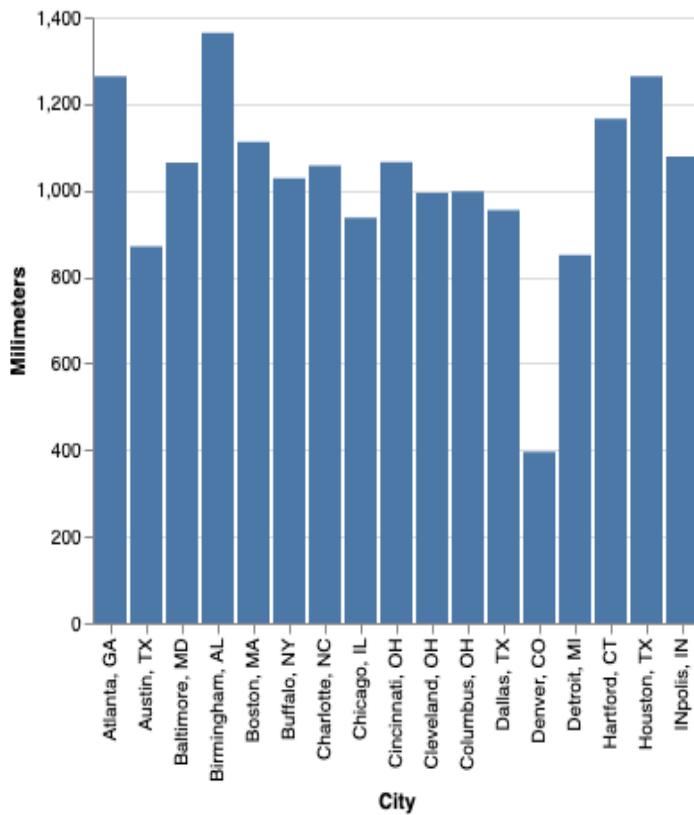
Mark: bar



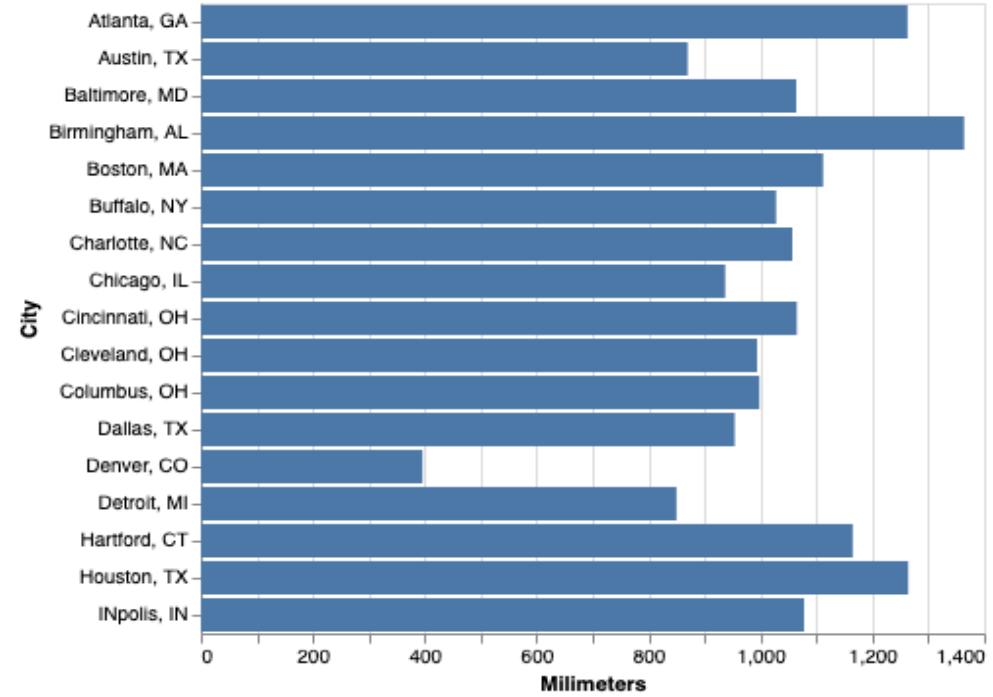
Toy data, bar charts

Mark: bar

Nominal (city name) => x
Quantitative (Annual precipitation) => y



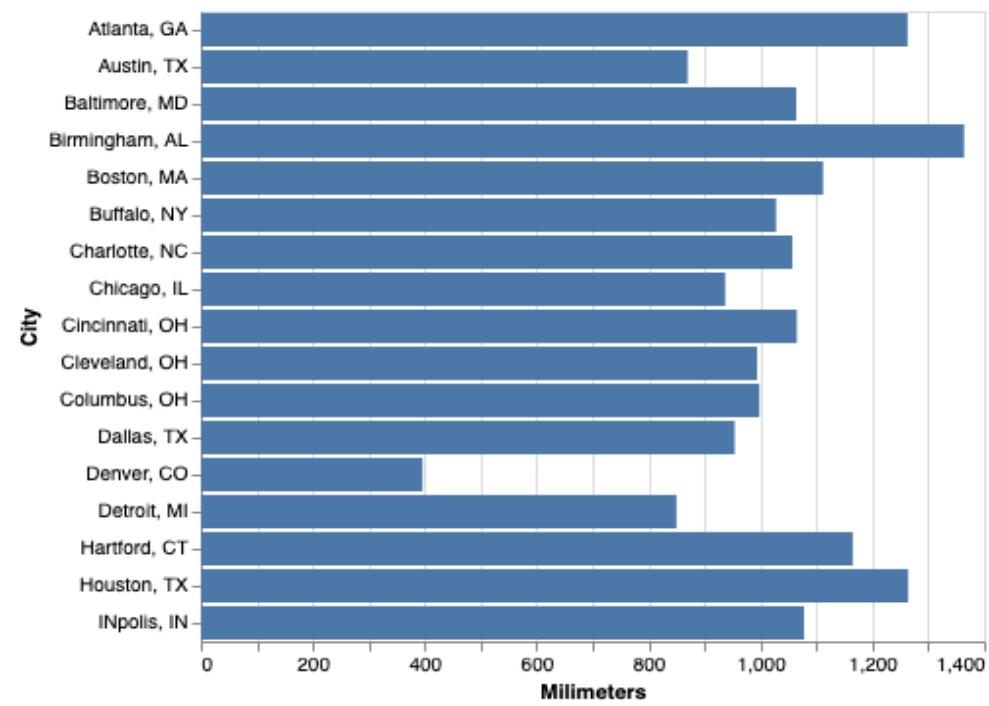
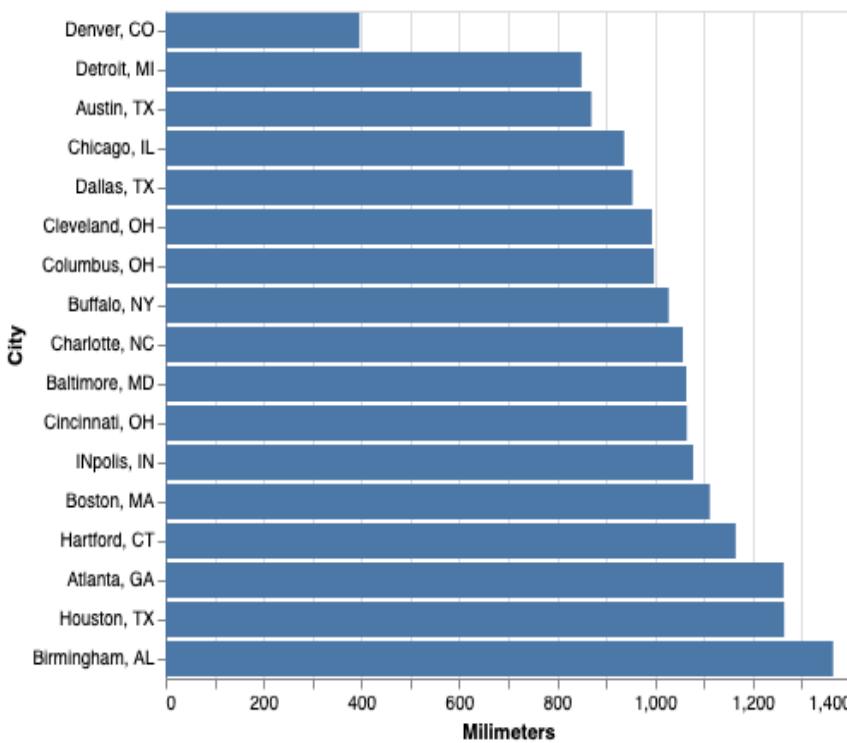
Nominal (city name) => y
Quantitative (Annual precipitation) => x



Toy data, bar charts

Nominal (city name) => y
Quantitative (Annual precipitation) => x

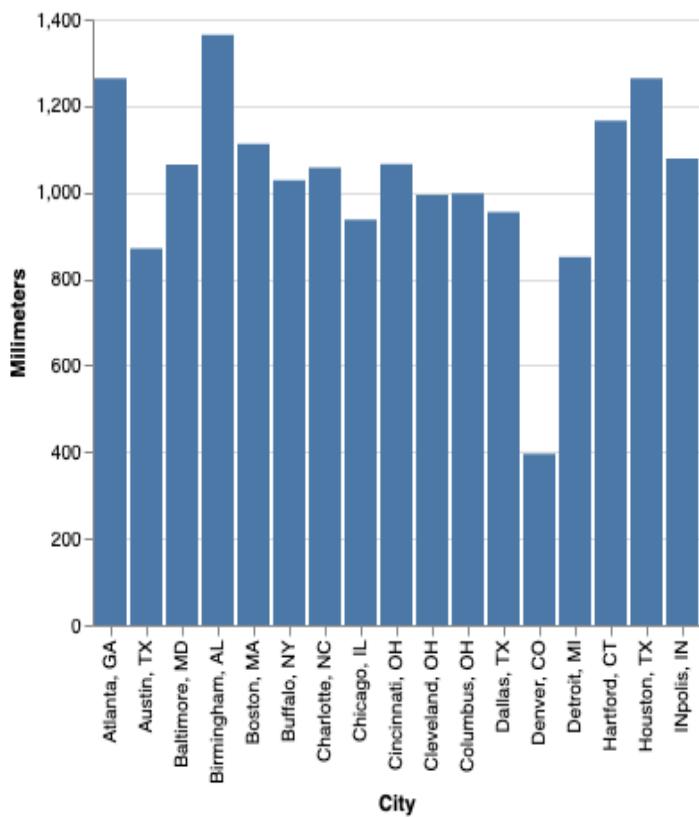
Mark: bar



Toy data, bar charts

Nominal (city name) => x
Quantitative (Annual precipitation) => y

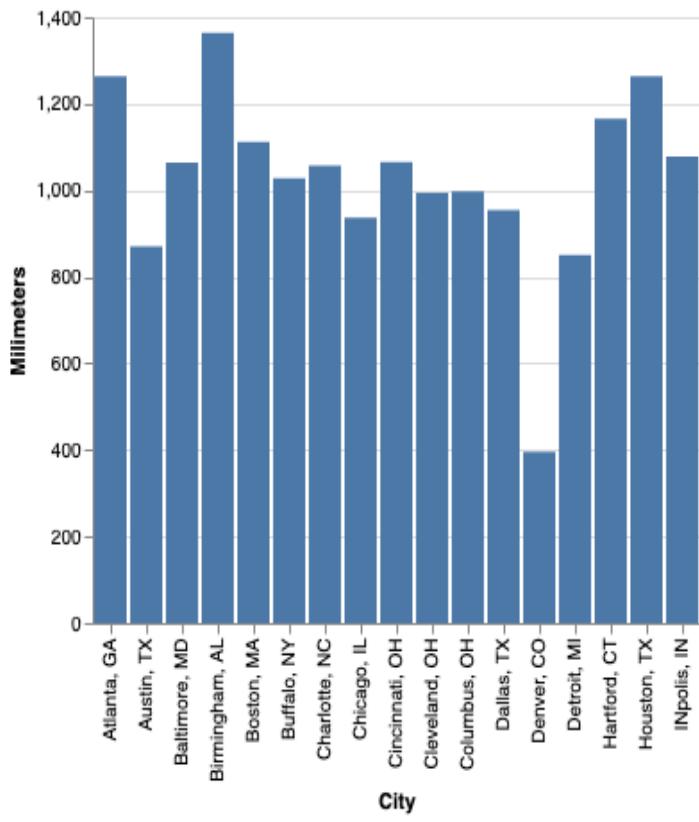
Mark: bar



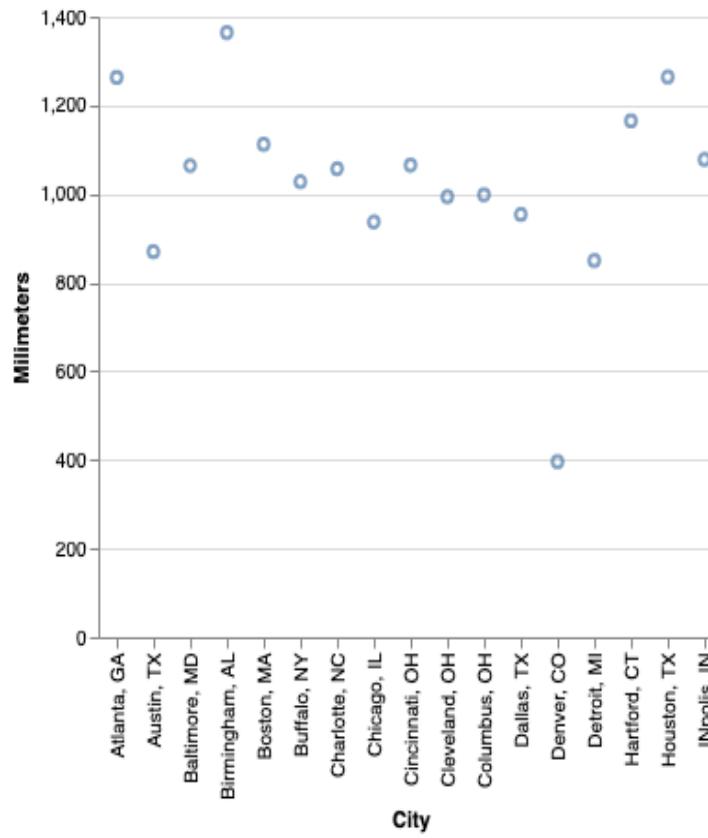
Mark types

Nominal (city name) => x
Quantitative (Annual precipitation) => y

Mark: bar



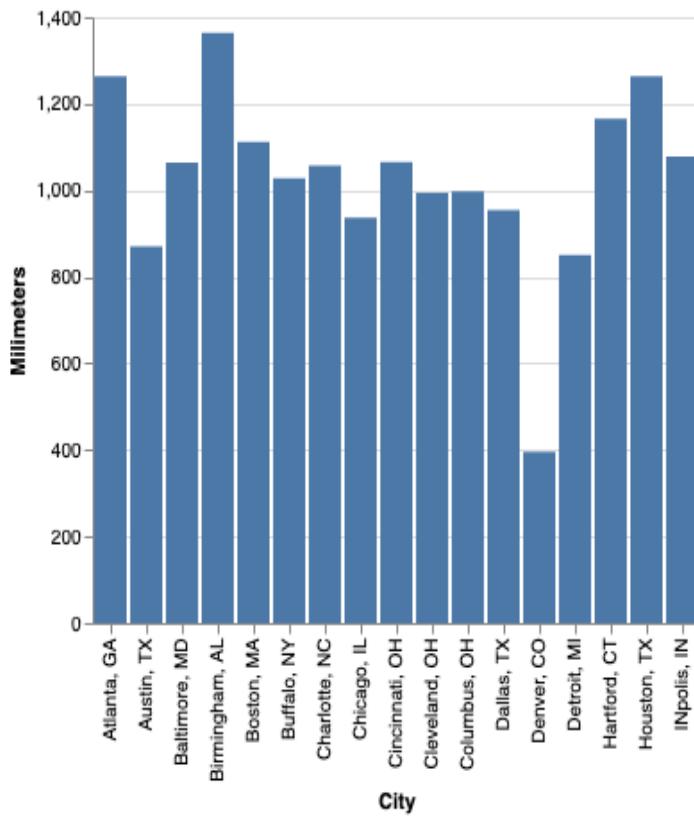
Mark: point



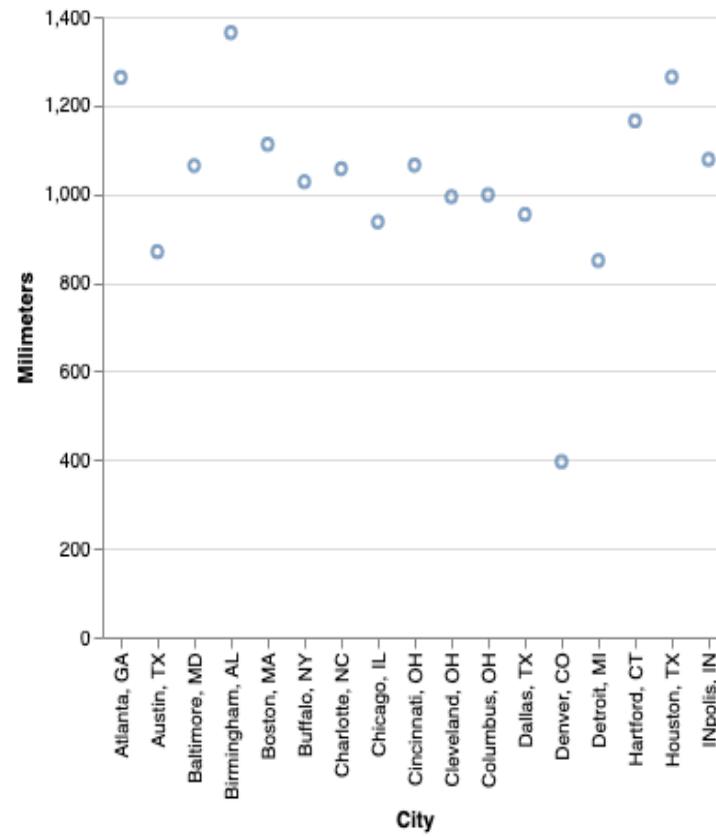
Mark types

Nominal (city name) => x
Quantitative (Annual precipitation) => y

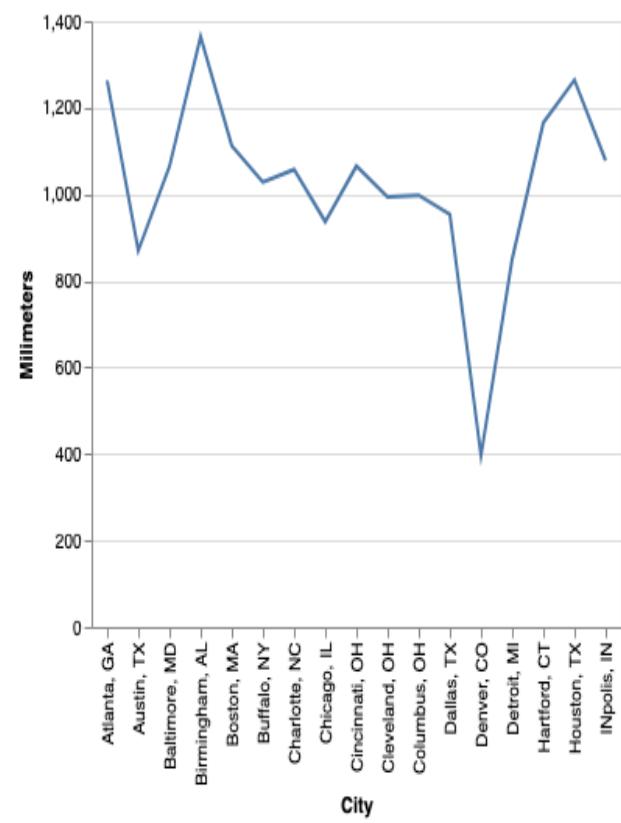
Mark: bar



Mark: point



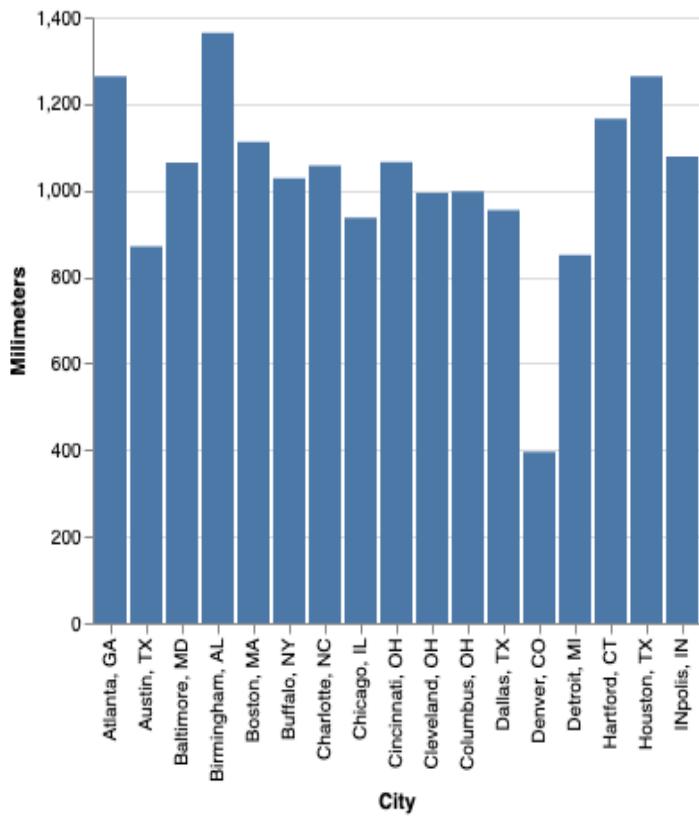
Mark: line



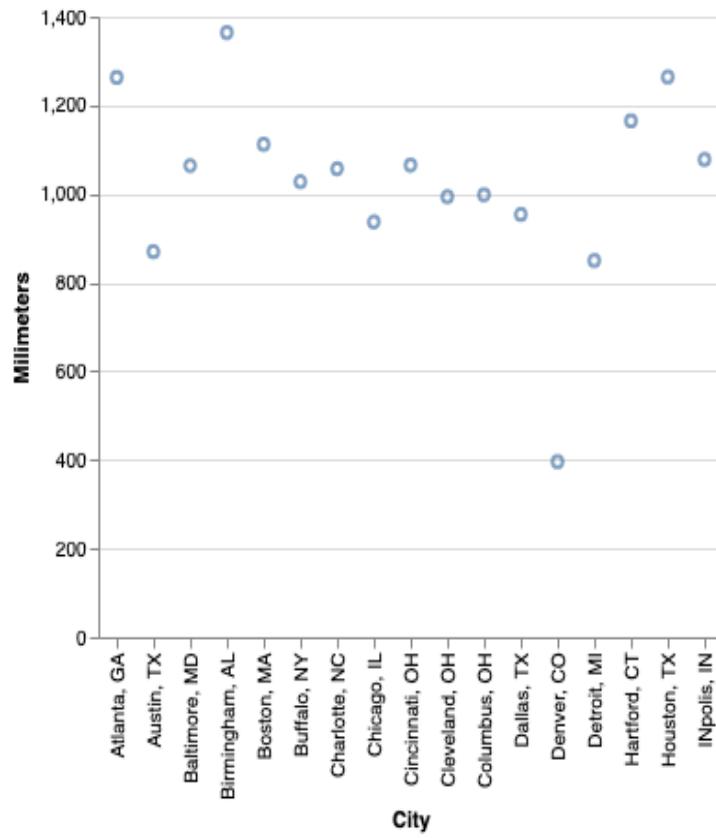
Mark types

Nominal (city name) => x
Quantitative (Annual precipitation) => y

Mark: bar



Mark: point

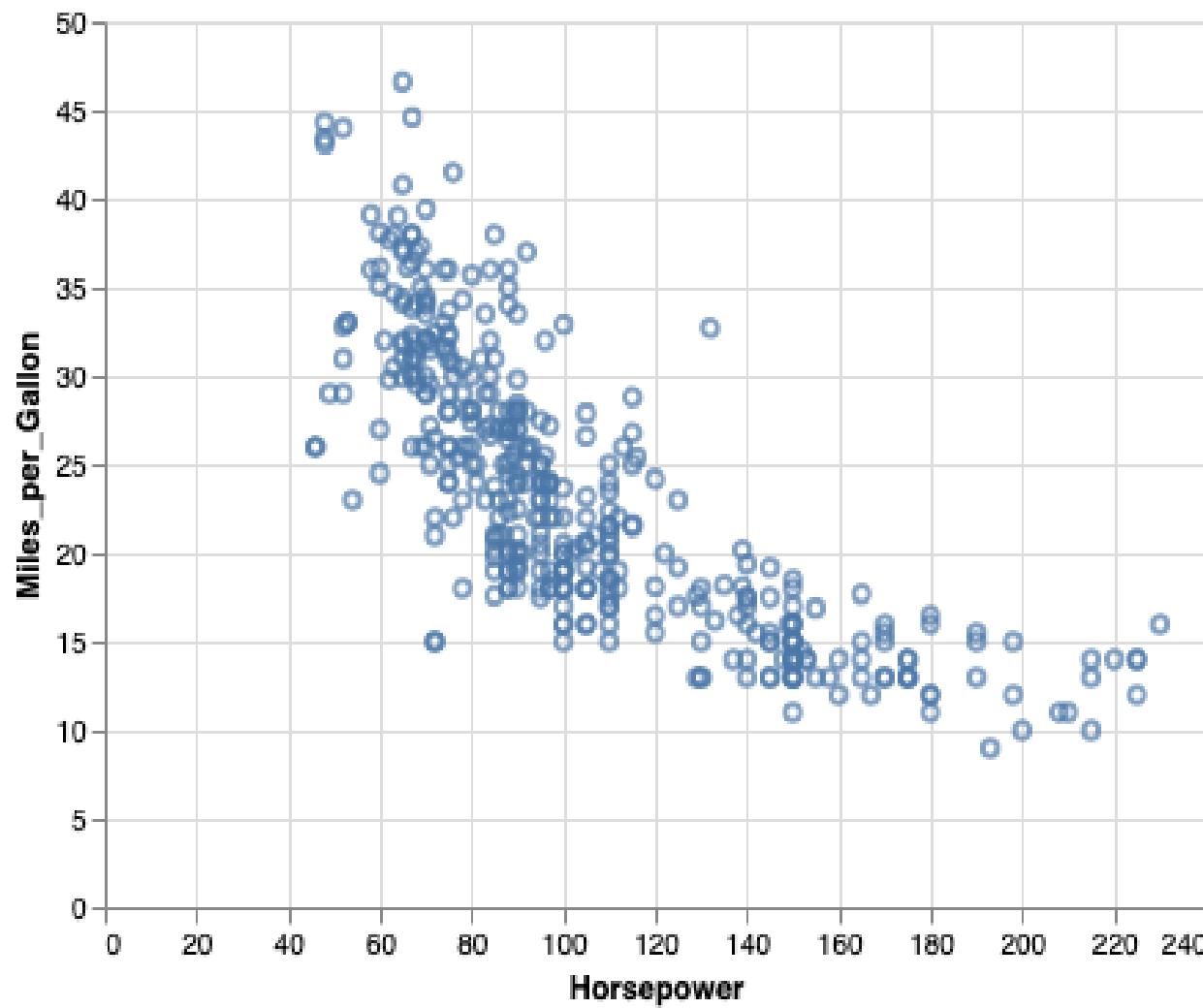


Mark: line



Scatterplot

Quantitative (Horsepower) => x
Quantitative (Miles_per_Gallon) => y



Scatterplot with nominal color

Nominal (city name) => x
Quantitative (Annual precipitation) => y
Nominal (Origin) => color

