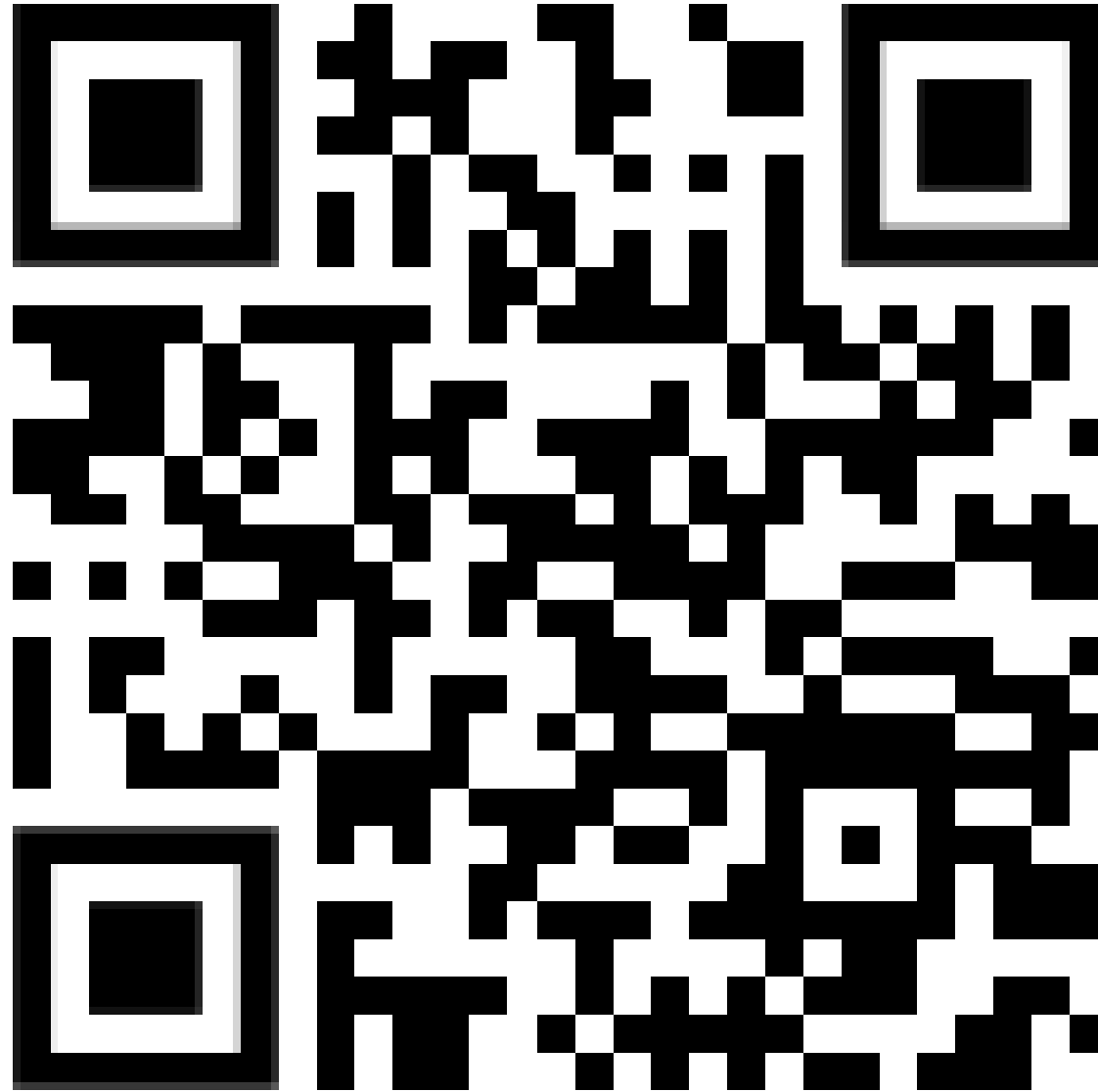




# CS-GY 6313 B: Information Visualization

9/12/2024

# Programming experience



[tinyurl.com/f6sby8su](https://tinyurl.com/f6sby8su)

# Example Final Projects

- <https://drive.google.com/file/d/1VMluv29u4wwjghJalscGvCDYK0XALxPD/view>
- <https://mbtaviz.github.io/>
- <https://www.washingtonpost.com/graphics/2020/world/corona-simulator/>
- Your implementation of an existing research paper



# Data Types

# Data types

- Need to process data with a computer. What does it look like? How can we organize it?

# Data types

- 5 basic data types:
  - Items
  - Attributes
  - Links
  - Positions
  - Grids
- Provides a “language” that we can use to organize data viz concepts

# Dataset types

- Collection of data → dataset
- Can be static or dynamic

## Tables

- Items
- Attributes

## Networks

- Items (nodes)
- Links
- Attributes

## Fields

- Grids
- Positions
- Attributes

## Geometry

- Items
- Positions
- Attributes (sometimes)

# Dataset types: Tables

- Much of the data we consider in life can be arranged in a table

Country	Year	Population	Case
A	2020	10	1
A	2021	11	2
B	2020	50	3
B	2021	55	5



# Dataset types: Tables

- Each **variable** must have its own **column**
- Each **observation** must have its own **row**
- Each **value** must have its own **cell**

## Rule of thumb:

- Variables often map to aesthetics
- Observations can be compared

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	214258	1272915272
China	2000	218766	128042583

variables

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	214258	1272915272
China	2000	218766	128042583

observations

country	year	cases	population
Afghanistan	99	745	19987071
Afghanistan	00	2666	20595360
Brazil	99	37737	172006362
Brazil	00	80488	174504898
China	99	214258	1272915272
China	00	218766	128042583

values

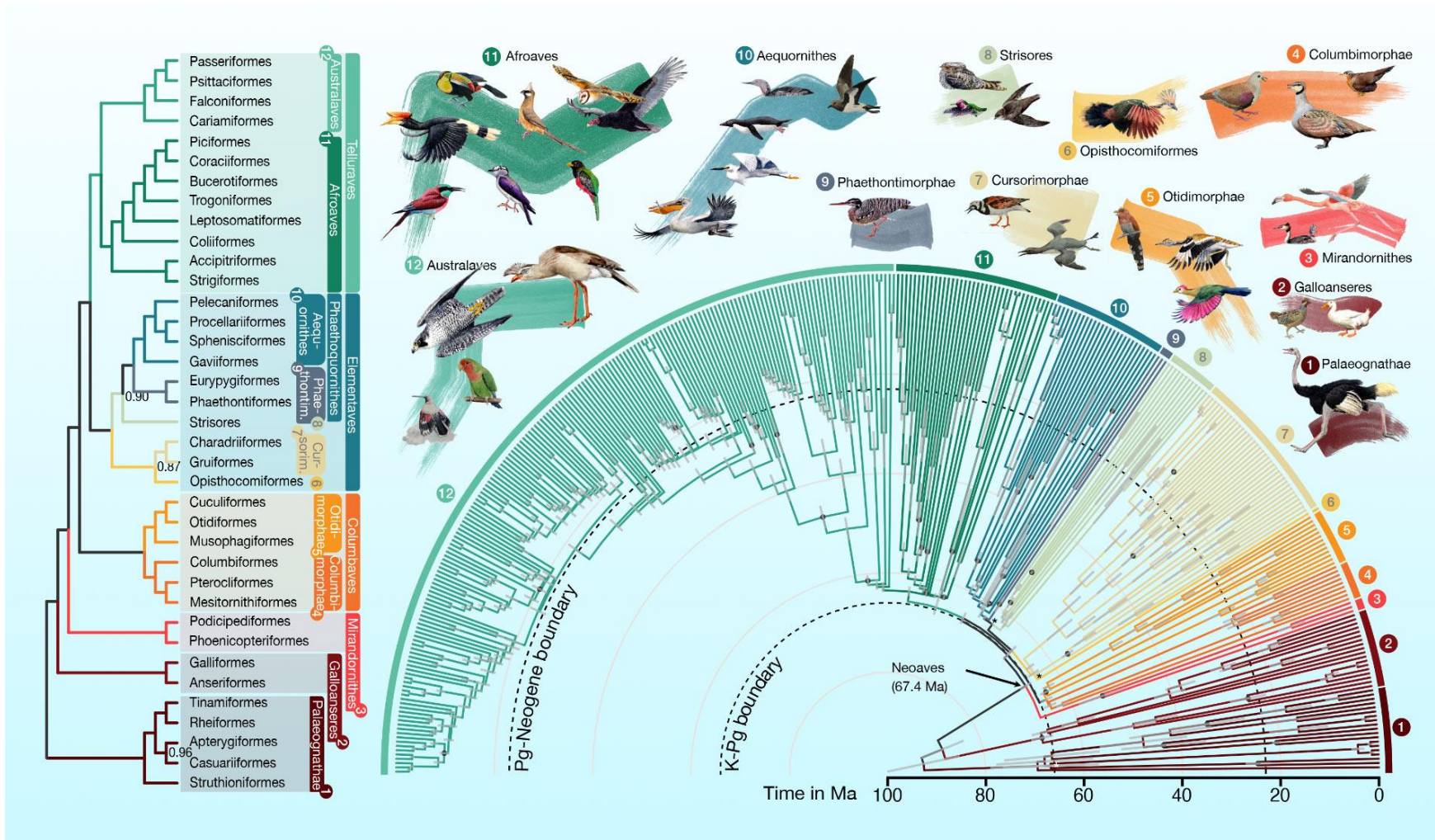
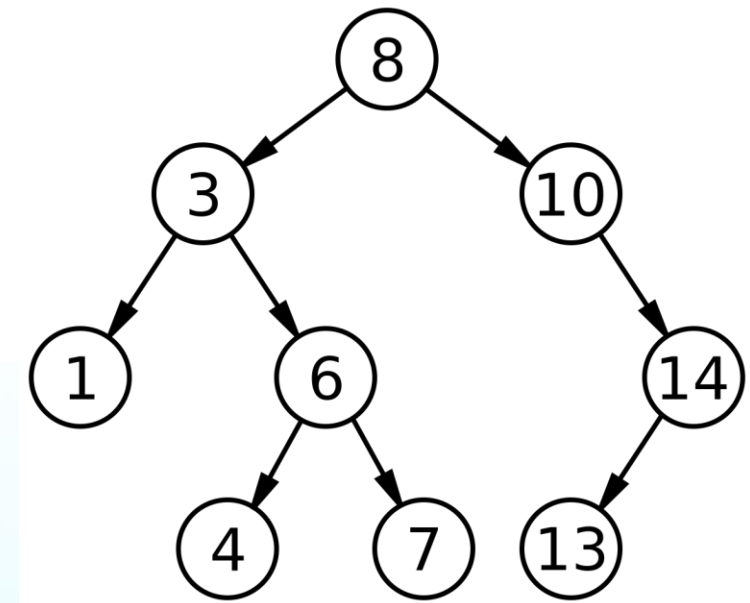
# Dataset types: Networks

- Good for specifying relationships between items
- A **node** is an item in a network (vertex)
- A **link** is a relation between two items (edge)
- Nodes and links often have attributes
  - Do not have to be related
- Examples:
  - Social network
  - Computer network
  - Gene interaction network



# Dataset types: Networks - Trees

- A tree is a special type of network
- Trees have **hierarchical** structure





# Dataset types: Fields

- Each cell is a measurement/calculation from a **continuous** domain
  - Infinite number of possible measurements, but fields discretize the domain
- **Sampling strategy** is important
  - Don't want to mislead the reader
  - Interpolation can help (can be misleading!)

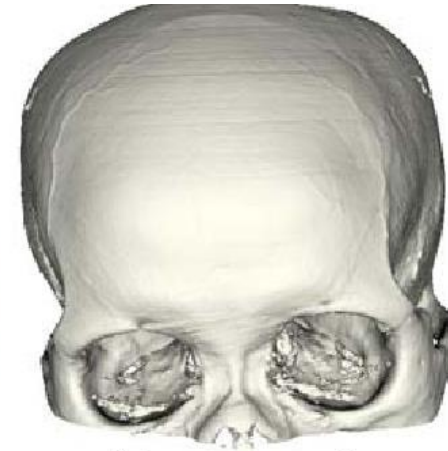


# Dataset types: Fields

- Each cell is a measurement/calculation from a **continuous** domain
  - Infinite number of possible measurements, but fields discretize the domain
- **Sampling strategy** is important
  - Don't want to mislead the reader
  - Interpolation can help (can be misleading!)
- Examples:
  - Medical scans



(a) skin reconstruction



(b) bone reconstruction



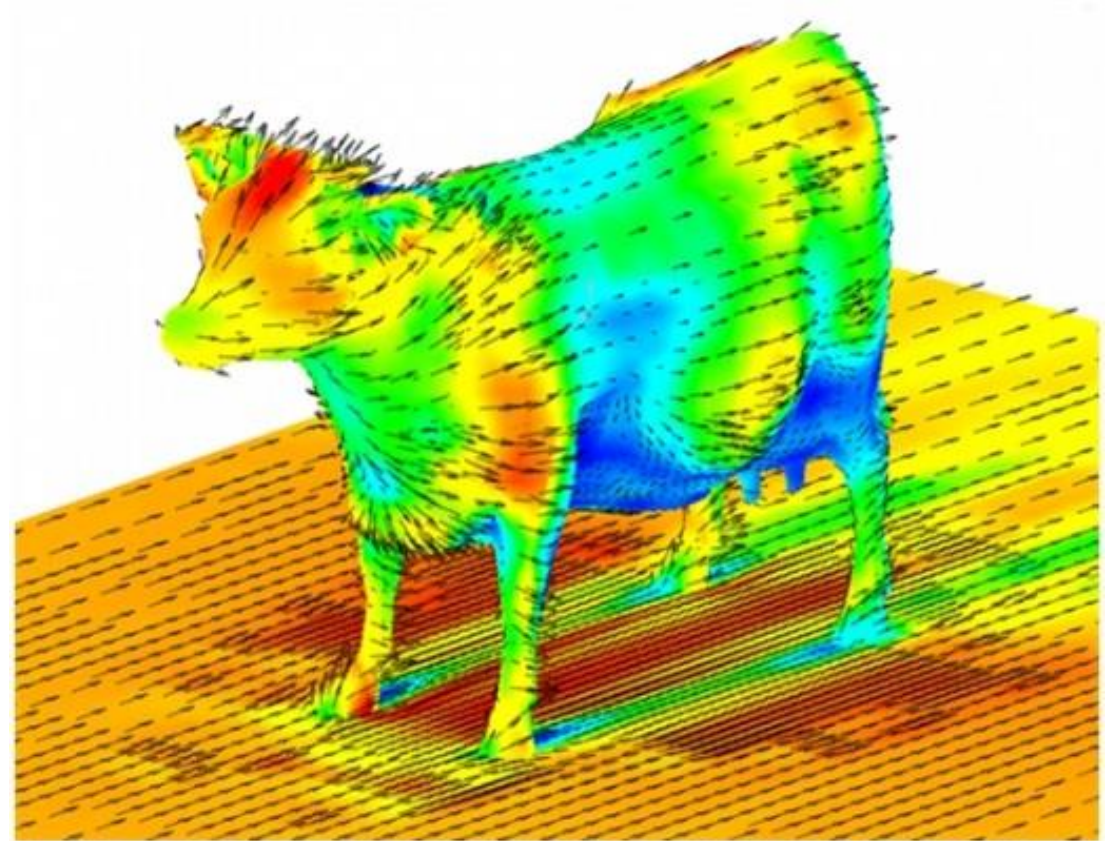
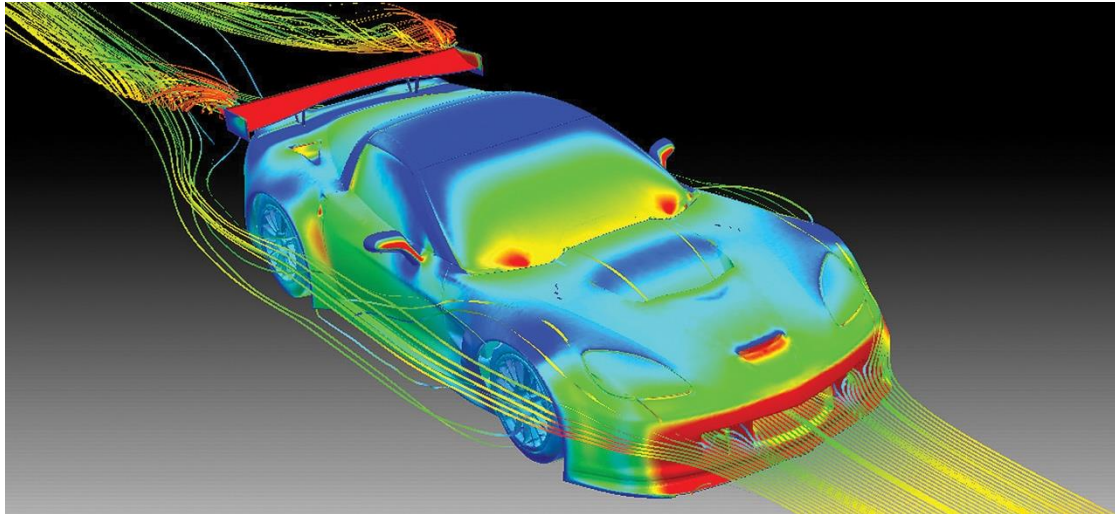
(c) the skin, bone reconstruction



(d) the skin, bone reconstruction :

# Dataset types: Fields

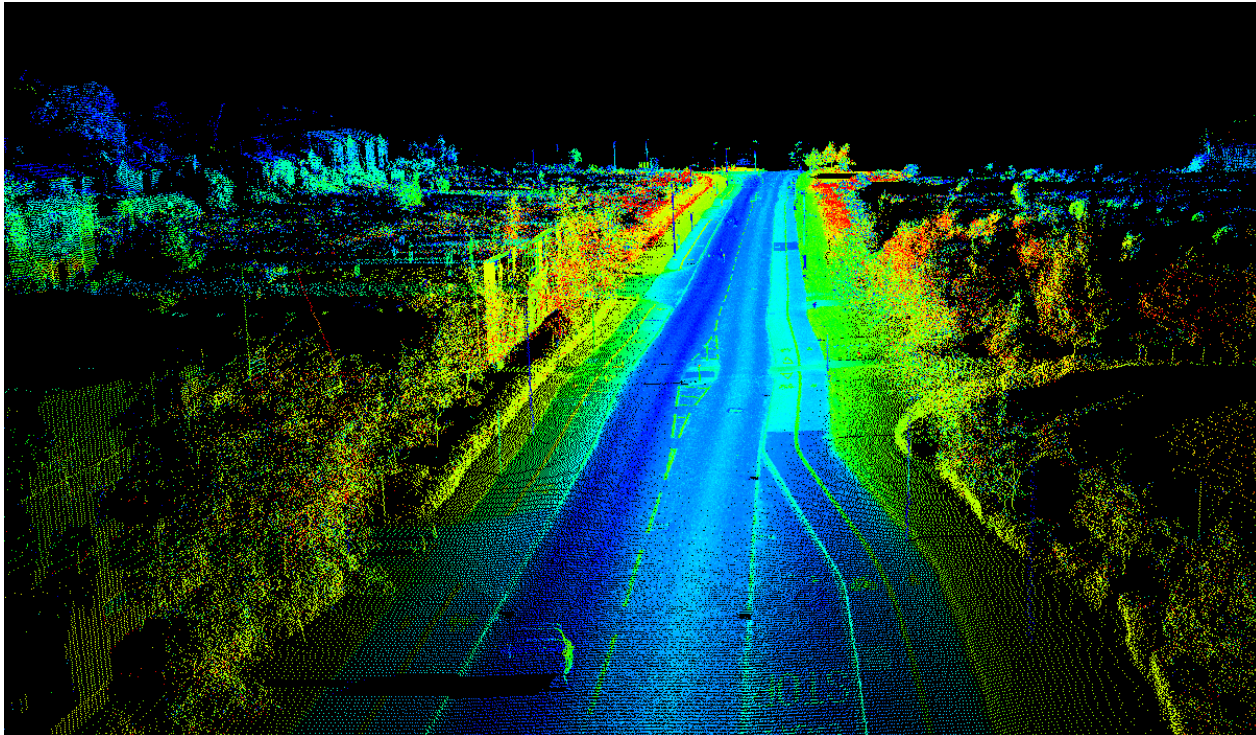
- Each cell is a measurement/calculation from a **continuous** domain
  - Infinite number of possible measurements, but fields discretize the domain
- **Sampling strategy** is important
  - Don't want to mislead the reader
  - Interpolation can help (can be misleading!)
- Examples:
  - Medical scans
  - Aerodynamics data





# Dataset types: Geometry

- Specifies info about the shape of items
  - Explicit spatial positions
- Intrinsically spatial (like fields)
- May or may not have attributes!
  - Examples: Roads, Lidar scans



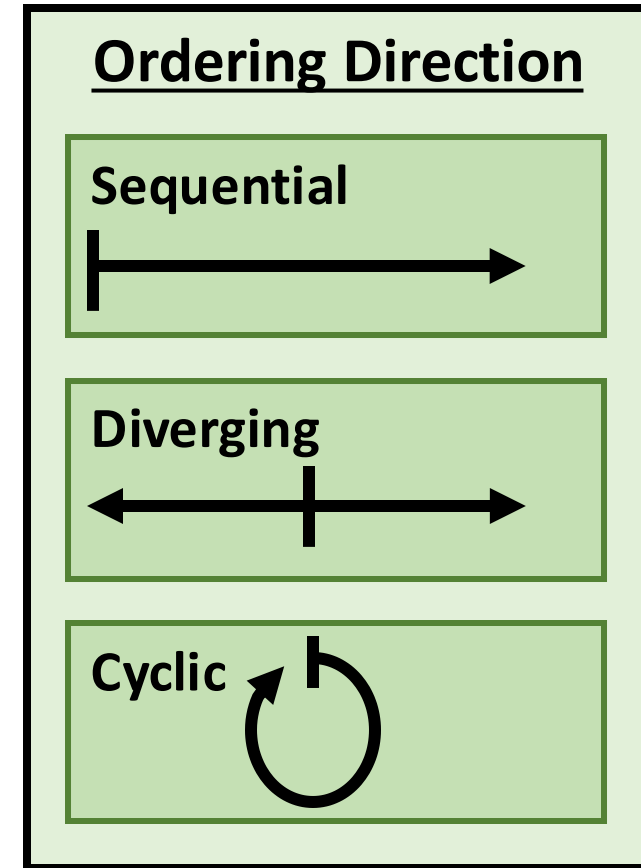
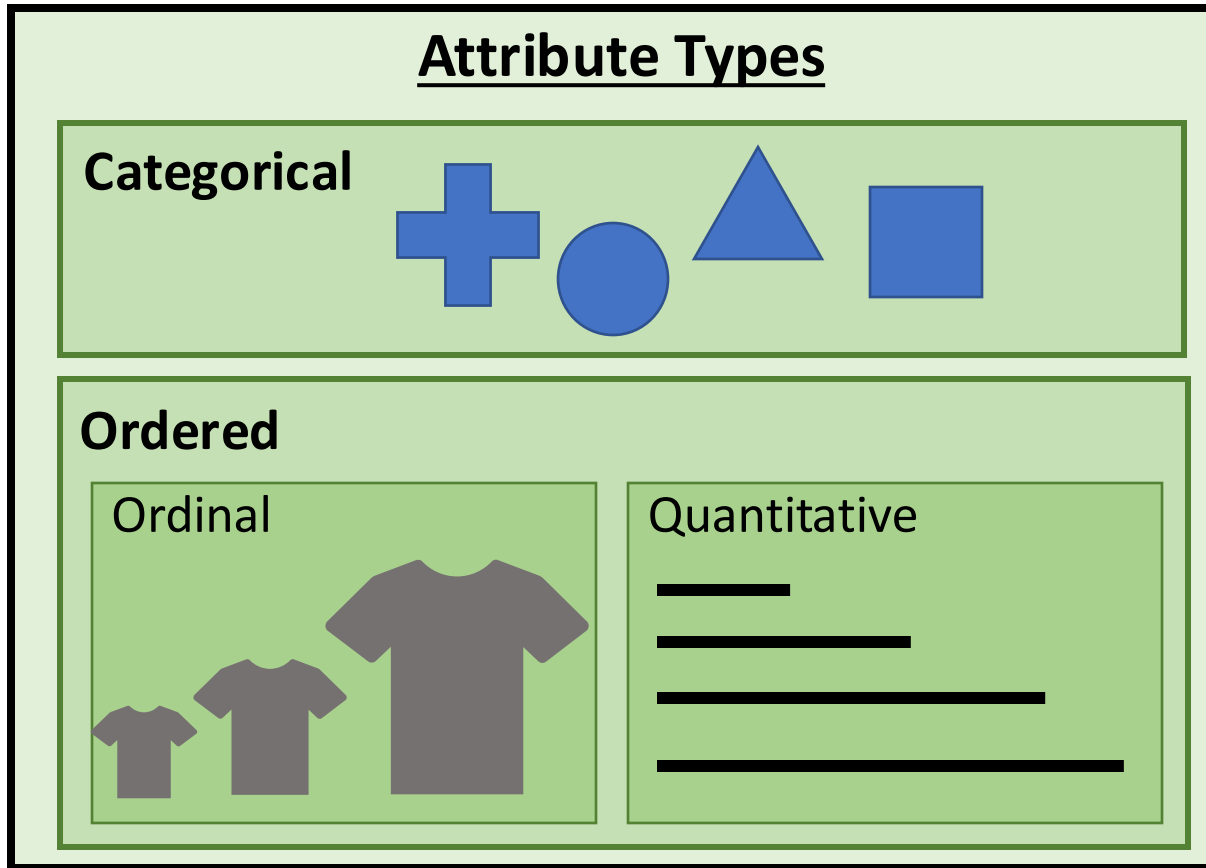
# Attribute types

- Attributes are the measured properties of items
  - Person's height
  - Density of a cloud
  - Annual rainfall for a country
- Usually maps onto visual aesthetics of the viz



# Attribute types

- Different kinds of attributes



# Attribute types: Categorical (nominal)

- Does not have an implicit ordering
- Also called **nominal** data
- Can still be a number, but doesn't make sense to aggregate
- Examples:
  - Dog breeds (hierarchical!)
  - Song genres
  - City names

# Attribute types: Ordered

- Ordered data has an implicit ordering
- **Ordinal**
  - Has an order but differences between categories may not be equal/measurable
  - Example:
    - Likert scale (questionnaire responses)
- **Quantitative**
  - Supports arithmetic, differences between values are measurable
  - **Interval data:** We can derive the gap between data values, and it has meaning
  - **Ratio data:** Similar to interval data, but has a true zero that has meaning. Allows math operations
  - Example:
    - 1, 2, 3, 4, 5.4, 345.2221, 17366105.23872, 1000000000000

# Attribute types: Ordered

- Data may have different types of ordering
- **Sequential:** Values have a minimum and a maximum
  - Example: Person's height
- **Diverging:** Values can be divided into 2 sequences pointing in opposite directions, meeting at a common zero point
  - Example: Terrain elevation (sea level = 0)
- **Cyclic:** Values wrap around back to a starting point
  - Example: Hour of the day

# Attribute hierarchies

- Attributes can often be grouped hierarchically
- Interesting patterns may emerge at different hierarchy levels
- Example: stock price



# Attribute types

- Also makes sense to group into **dependent** and **independent**
- **Independent attributes:**
  - Way of describing the data
  - Usually discrete (i.e., categorical)
- **Dependent attributes:**
  - Their value is a function of one or more independent variables
  - Numerical data that can be analyzed, manipulated, and aggregated

# Metadata

- Data about data
- Can be difficult to visualize since it adds complexity
  - Underlying task of visualization remains the same, though

# Exercise: US Census

- What are the types of these attributes?
  - Categorical, ordinal, or quantitative (ratio or interval)?
  - Dependent or independent?
- Talk with your neighbor for ~2 min
- **People count:** # of people in group
- **Year:** 1850 – 2000 (every decade)
- **Age:** 0 – 90+
- **Sex:** male, female
- **Marital status:** single, married, divorced

	A	B	C	D	E
1	year	age	marst	sex	people
2	1850	0	0	1	1483789
3	1850	0	0	2	1450376
4	1850	5	0	1	1411067
5	1850	5	0	2	1359668
6	1850	10	0	1	1260099
7	1850	10	0	2	1216114
8	1850	15	0	1	1077133
9	1850	15	0	2	1110619
10	1850	20	0	1	1017281
11	1850	20	0	2	1003841
12	1850	25	0	1	862547
13	1850	25	0	2	799482
14	1850	30	0	1	730638
15	1850	30	0	2	639636
16	1850	35	0	1	588487
17	1850	35	0	2	505012
18	1850	40	0	1	475911
19	1850	40	0	2	428185
20	1850	45	0	1	384211
21	1850	45	0	2	341254
22	1850	50	0	1	301340



# Exercise: US Census

- What are the types of these attributes?
  - Categorical, ordinal, or quantitative (ratio or interval)?
  - Dependent or independent?
- Talk with your neighbor for ~2 min
- **People count:** quantitative (ratio)
- **Year:** quantitative (interval) or ordinal
- **Age:** quantitative (ratio) or ordinal
- **Sex:** categorical
- **Marital status:** categorical

	A	B	C	D	E
1	year	age	marst	sex	people
2	1850	0	0	1	1483789
3	1850	0	0	2	1450376
4	1850	5	0	1	1411067
5	1850	5	0	2	1359668
6	1850	10	0	1	1260099
7	1850	10	0	2	1216114
8	1850	15	0	1	1077133
9	1850	15	0	2	1110619
10	1850	20	0	1	1017281
11	1850	20	0	2	1003841
12	1850	25	0	1	862547
13	1850	25	0	2	799482
14	1850	30	0	1	730638
15	1850	30	0	2	639636
16	1850	35	0	1	588487
17	1850	35	0	2	505012
18	1850	40	0	1	475911
19	1850	40	0	2	428185
20	1850	45	0	1	384211
21	1850	45	0	2	341254
22	1850	50	0	1	221242

# Exercise: US Census

- What are the types of these attributes?
  - Categorical, ordinal, or quantitative (ratio or interval)?
  - Dependent or independent?
- Talk with your neighbor for ~2 min
- **People count:** dependent
- **Year:** independent
- **Age:** dependent (?)
- **Sex:** independent
- **Marital status:** independent (?)

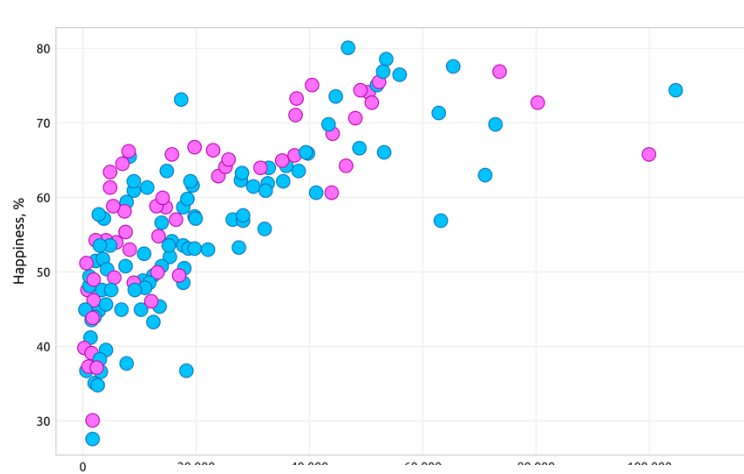
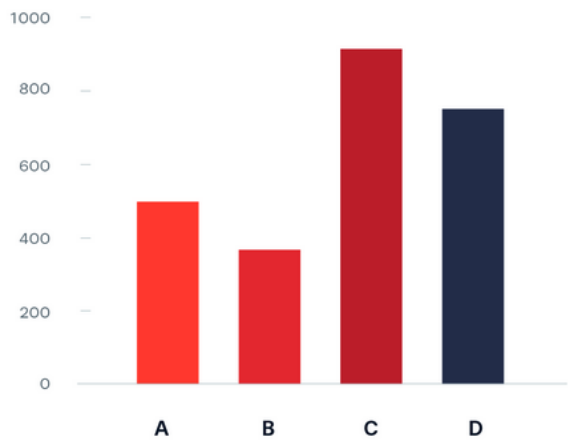
	A	B	C	D	E
1	year	age	marst	sex	people
2	1850	0	0	1	1483789
3	1850	0	0	2	1450376
4	1850	5	0	1	1411067
5	1850	5	0	2	1359668
6	1850	10	0	1	1260099
7	1850	10	0	2	1216114
8	1850	15	0	1	1077133
9	1850	15	0	2	1110619
10	1850	20	0	1	1017281
11	1850	20	0	2	1003841
12	1850	25	0	1	862547
13	1850	25	0	2	799482
14	1850	30	0	1	730638
15	1850	30	0	2	639636
16	1850	35	0	1	588487
17	1850	35	0	2	505012
18	1850	40	0	1	475911
19	1850	40	0	2	428185
20	1850	45	0	1	384211
21	1850	45	0	2	341254
22	1850	50	0	1	331345



5 Minute Break  
resume @ 11:58am

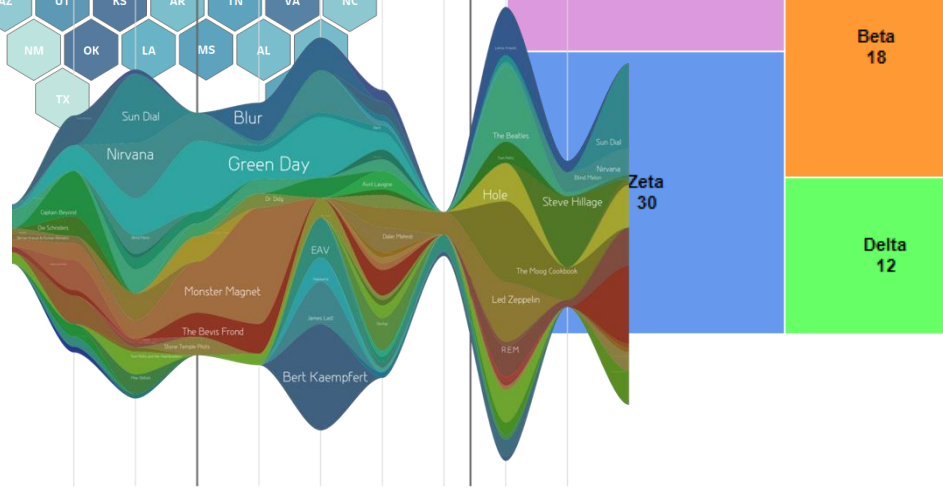
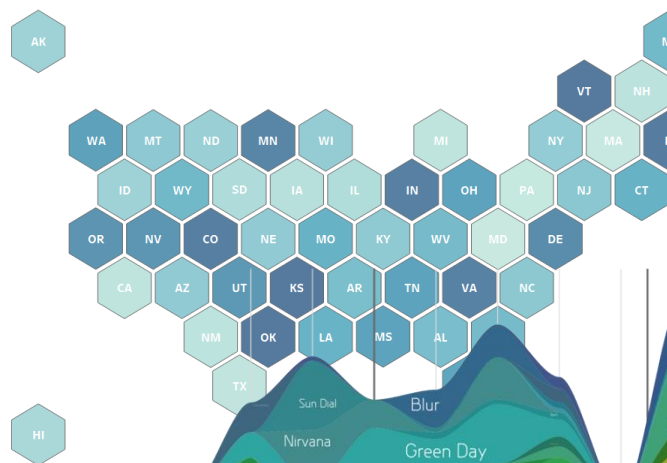
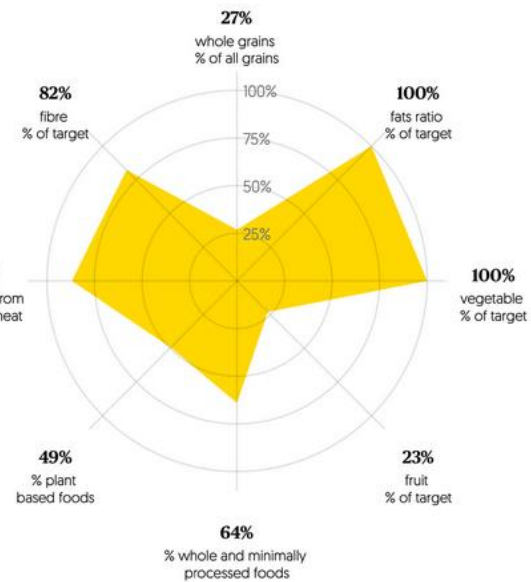
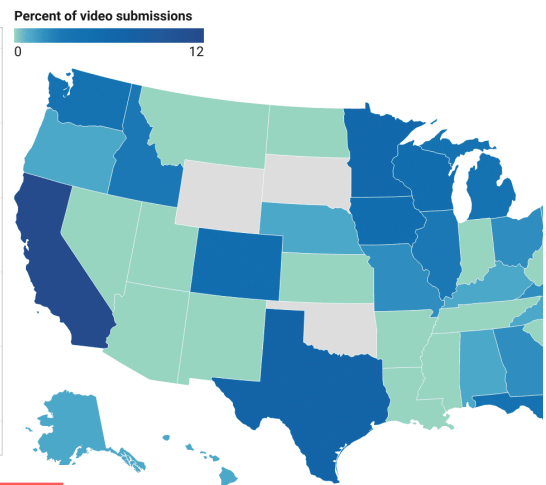


# Visualization techniques



Percentage of Video Submissions by State to the World of 7 Billion Student Video Competition 20

A choropleth map of the United States showing the percentage of videos sub in each state.

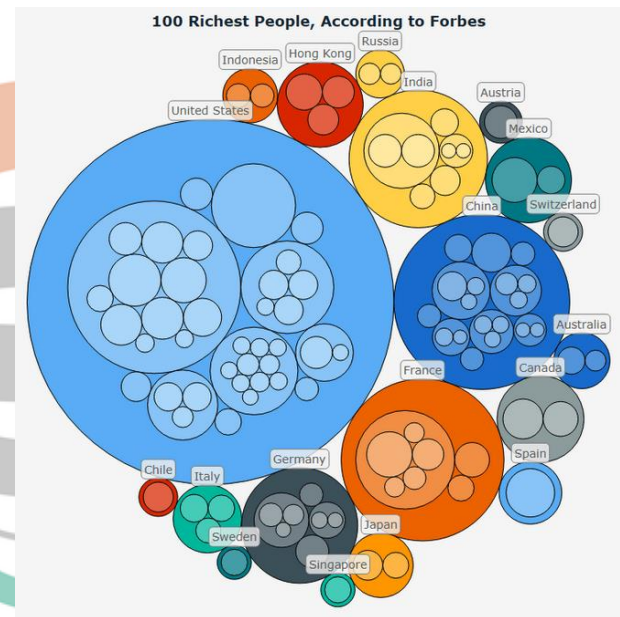


Datawrapper

Wages

Other

Budget



# Python + matplotlib

- [https://colab.research.google.com/drive/1PJqGp\\_v-cKx62hC0ba9g2zzxzJ8NySVq?usp=sharing](https://colab.research.google.com/drive/1PJqGp_v-cKx62hC0ba9g2zzxzJ8NySVq?usp=sharing)

# Python + bokeh

- <https://colab.research.google.com/drive/1SYAirRVHoGK-zxXyMn4VT0yONMEDiY8-?usp=sharing>

# LaTeX crash course

- Essentially a Markup Language (like HTML, XML and RTF)
- Invented in the 70s by Donald Knuth
  - He was writing a book, but there were no good typesetting solutions, so he made his own!
- You should use <https://www.overleaf.com/>
  - Requires internet connection, but avoids somewhat annoying installation + compile process

# LaTeX vs Word processors

- High typeset quality
- Easy to include math formulas
- Source file format is not bounded to a particular OS or platform
- LaTeX implementations exist for all platforms (DOS, Windows, Unix, ...)
- LaTeX is free



# LaTeX vs Word processors

- De facto standard for scientific publishing
- Very few bugs
- Good for large documents
- Not very easy to learn 😞

# LaTeX file structure

- Document Class
  - Predefined Formats (article, report, book,..)
- Packages used
  - Added Functionality (graphics, reference style,...)
- Main Body
  - Text and Bibliography References

# LaTeX basics

- Document class

```
\documentclass[options]{class}
```

options = a4paper, 11pt, 12pt, 10pt, twocolumn, landscape,...

class = article, report, book,...

- Packages

```
\usepackage{package name}
```

epsfig = insert PS pictures into the document

fancyhdr = easy definition of footer and header

# Body of text

- Start with `\begin{document}`
- End with `\end{document}`
- Typesetting Text
  - `\\` or `\newline` and `\newpage`
  - Quotations use backtick (```) and apostrophe (`'`)
  - Bold `\textbf{.....}` or `\bf`
  - Italics `\emph{.....}` or `\textit{.....}` or `\it`
  - Underline `\underline{.....}` or `\ul`

# Organizing a document

- Sections
  - `\section{Latex is Great}` = 1. Latex is Great
  - `\subsection{Why Latex is Great}` = 1.1 Why Latex is Great
  - `\subsubsection{Reason One}` = 1.1.1 Reason One
  - `\appendix` - changes numbering scheme
  - `\chapter{...}` - To be used with book and report document classes
- Titles, authors, and others
  - `\title{...}`
  - `\author{...}`
  - `\footnote{...}`
- Labels
  - `\label{marker}` - Marker in document.
  - `\ref{marker}` - Displays section location of marker.

# Lists

- LaTeX source:

```
\begin{itemize}  
  \item Apple  
  \item Orange  
\end{itemize}
```

- Result:

- Apple
- Orange

- `Enumerate` instead of `itemize` gives a numbered list

- Lists can be nested

- `\begin{...}` is called an *environment* in LaTeX

- Itemize, figure, table, enumerate, etc

# Example document

```
\documentclass{article}  
\title{Simple Example}  
\author{Andrei Gurtov}  
\date{March 2000}  
\begin{document}  
\maketitle  
Hello world!  
\end{document}
```

Simple Example

Andrei Gurtov

March 2000

Hello world!

# Tables

```
\begin{tabular}{|l|r|c} \hline
Date & Price & Size \\ \hline
Yesterday & 5 & big \\ \hline
Today & 3 & small \\ \hline
\end{tabular}
```

Date	Price	Size
Yesterday	5	Big
Today	3	Small



# Figures

```
\usepackage{graphicx}
\graphicspath{ {./images/} }
\begin{document}
```

```
The universe is immense and it
seems to be homogeneous, in a large
scale, everywhere we look at.
```

```
\includegraphics{universe}
```

```
There's a picture of a galaxy above
```

```
\end{document}
```

The universe is immense and it seems to be homogeneous, in a large scale,  
everywhere we look at.



There's a picture of a galaxy above

# Bibliography

- Bibliography information is stored in a \*.bib file, in Bibtex format.
- Include chicago package
  - `\usepackage{chicago}`
- Set referencing style
  - `\bibliographystyle{chicago}`
- Create reference section by
  - `\bibliography{bibfile with no extension}`

# Bibliography

- Citing references in text
  - `\cite{cuc98}` = (Cuce 1998)
  - `\citeN{cru98}` = Crud (1998)
  - `\shortcite{tom98}` = (Tom, et. al. 1998)
- Creating Bibtex Files

Google Scholar

evolution of cats

Articles About 1,420,000 results (0.09 sec)

Any time

Since 2024

Since 2023

Since 2020

Custom range...

[The evolution of cats](#)  
SJ O'Brien, WE Johnson, Scientific American, 297(1), 68-75

... ■ The family history of the **cat** has been resolved

... resolved family tree for **cats**. ■ 1

☆ Save [Cite](#) Cited by 170

×

Cite

MLA O'Brien, Stephen J., and Warren E. Johnson. "The evolution of cats." *Scientific American* 297.1 (2007): 68-75.

APA O'Brien, S. J., & Johnson, W. E. (2007). The evolution of cats. *Scientific American*, 297(1), 68-75.

Chicago O'Brien, Stephen J., and Warren E. Johnson. "The evolution of cats." *Scientific American* 297, no. 1 (2007): 68-75.

Harvard O'Brien, S.J. and Johnson, W.E., 2007. The evolution of cats. *Scientific American*, 297(1), pp.68-75.

Vancouver O'Brien SJ, Johnson WE. The evolution of cats. *Scientific American*. 2007 Jul 1;297(1):68-75.

[BibTeX](#) [EndNote](#) [RefMan](#) [RefWorks](#)

```
@article{o2007evolution,
  title={The evolution of cats},
  author={O'Brien, Stephen J and Johnson, Warren E},
  journal={Scientific American},
  volume={297},
  number={1},
  pages={68--75},
  year={2007},
  publisher={JSTOR}
}
```