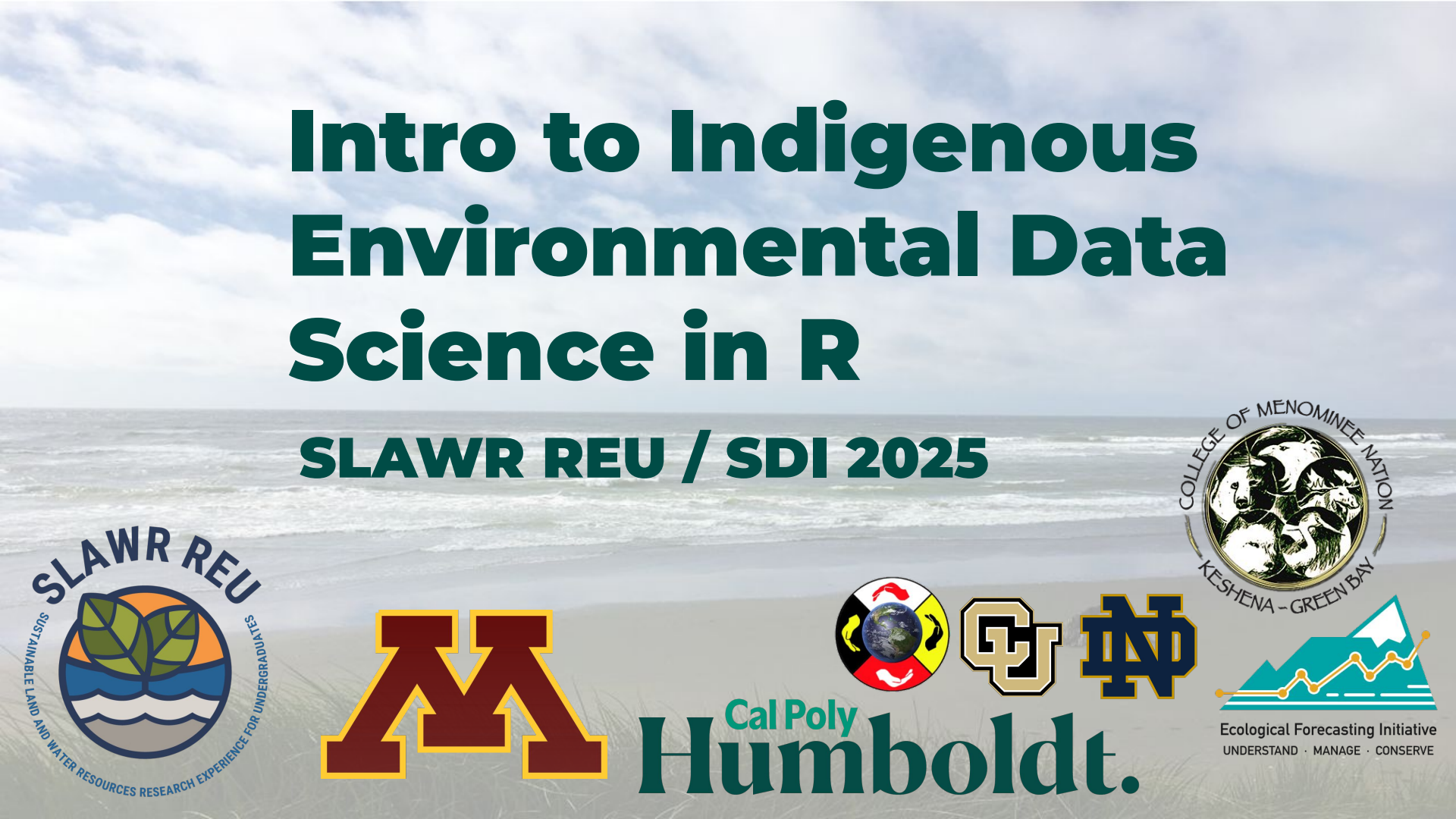
Intro to Indigenous Environmental Data Science in R

SLAWR REU / SDI 2025



Agenda

- Morning
  - Design Justice Principles, indigenous data sovereignty
  - intro to R code
- 12:00 PM Lunch and discussion
- 1:00 PM Microcredentials with Timberley
- Afternoon:
  - data visualization
  - intro to ggplot

telling someone's story with a graph

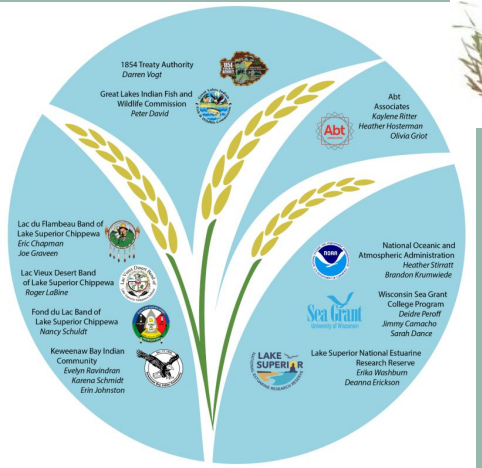
What is our responsibility to:

1. Our community, as the storytellers/designers?
2. The being whose story we are telling?
3. The community we are telling the story to?

telling someone's story

Who is our community? Whose story are we telling? Telling it to whom?

Lake Superior Manoomin Cultural and Ecosystem Characterization Study



The primary audiences for this report are Indigenous communities, tribal and non-tribal governments, and organizations who are working to actively manage and restore Manoomin across the Great Lakes.

- 2020 report, p. 2

telling someone's story with a graph

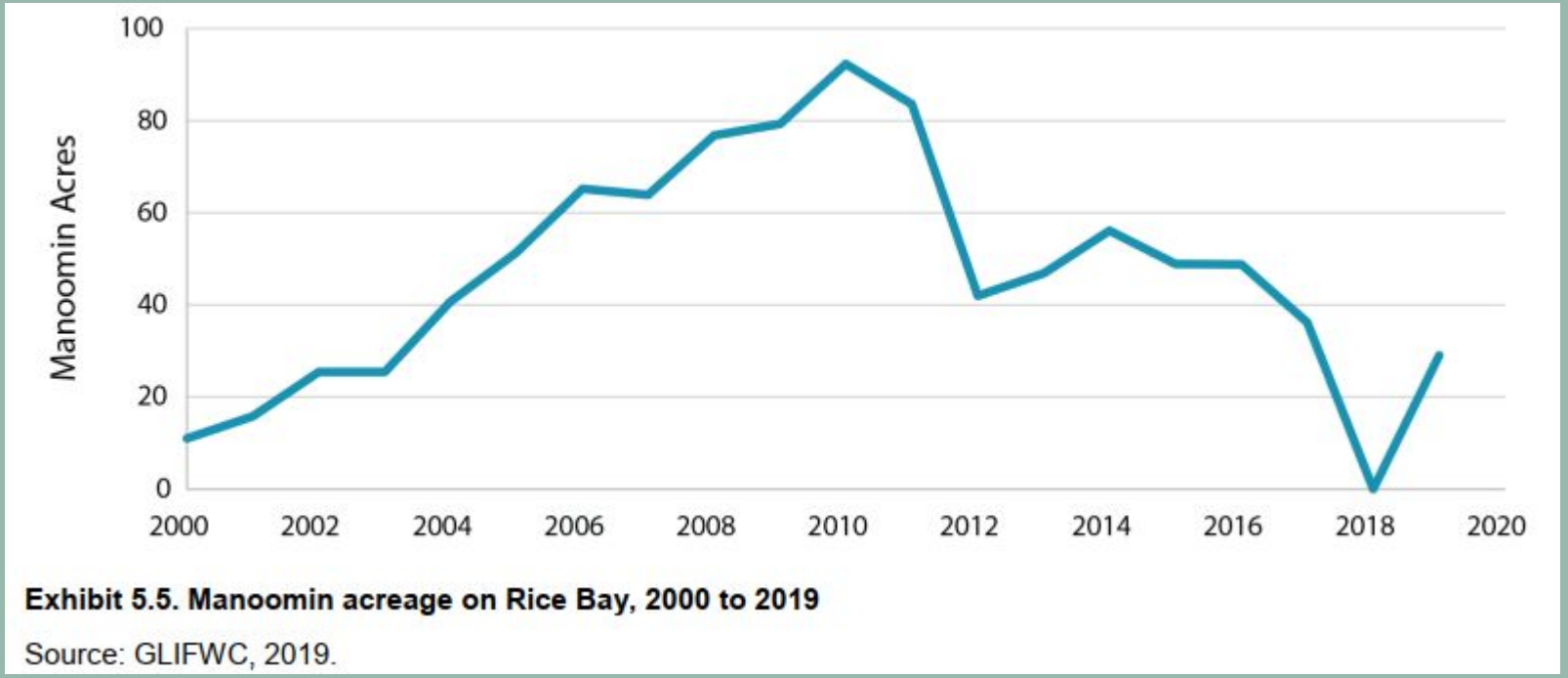


Exhibit 5.5. Manoomin acreage on Rice Bay, 2000 to 2019

Source: GLIFWC, 2019.

telling someone's story with a graph

12 things manoomin does.

Cultural Metrics

Anishinaabe (original people) – The place provides manoomin, which is sacred to the Anishinaabe and central to the foundations of their culture, sovereignty, and treaty rights.

Community relationships – Manoomin at this place contributes to bonding, traditions, and strengthening family and community connections.

Spirit relationships – Manoomin at this place enables the Anishinaabe to maintain connections and balance with spirit beings (or relatives) from all other orders of creation (first order: rock, water, fire and wind; second order: other plant beings; third order: animal beings; fourth order: human beings).

Manoominikewin – This place allows for the Anishinaabe to harvest, prepare, and share (gifting, healing, and eating) manoomin in the ways practiced by their ancestors for centuries.

Food sovereignty and health – This place provides the capacity to provide for the sustenance, health, and independence of the Anishinaabe.

Ecological Metrics

Biodiversity – Healthy manoomin and appropriate habitat at this place supports diverse biological communities (e.g., free of invasive species) that indicate the capacity of the place to support abundant associated plant and animal species (e.g., other native aquatic vegetation, fish, waterfowl, muskrat), providing for spiritual and subsistence needs.

Integrity – Physical habitat and hydrology, water and sediment chemistry support stands of manoomin that exhibit natural annual variability; viable seed bank ensures that sustainable manoomin populations will persist even after occasional poor production years. Natural genetic diversity is maintained without impact from cultivated strains, or reduced gene flow from the loss of nearby manoomin populations.

Water quality – This place has clean water (e.g., sulfate levels below 10 ppm) and sediments that can support robust stand density and wildlife diversity; is free of contamination or impacts from industrial, agricultural, recreational, or residential influence; and is of sufficient areal extent to sustain a manoomin population.

Water level – This place has a natural or managed hydrologic regime that can maximize resilience under variable or extreme climatic conditions across the growing season (maintaining optimal depth range and flow).

Cultural and Ecological Education Metrics

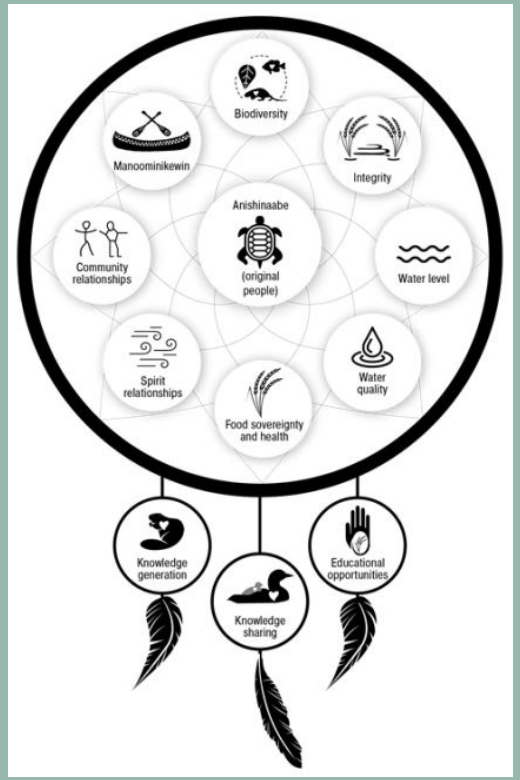
Knowledge generation – This place allows for continued learning and generation of the Anishinaabe practices, values, beliefs, and language through experience.

Knowledge sharing – This place allows for the continued sharing and transmittal of the Anishinaabe practices, values, beliefs, and language among family members and community.

Educational opportunities – This place provides opportunities for language, land stewardship, and other educational programs, such as educational rice camps.

telling someone's story with a graph

12 things manoomin does.



How they changed over 100 years.

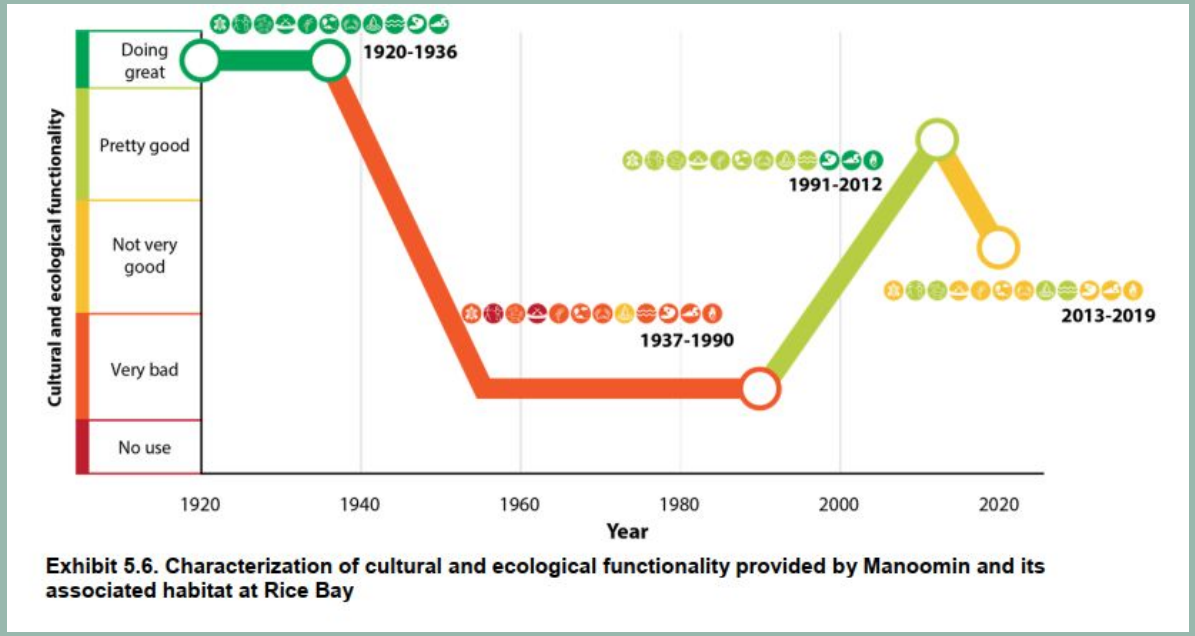


Exhibit 5.6. Characterization of cultural and ecological functionality provided by Manoomin and its associated habitat at Rice Bay

design justice principles

Principle 1

We use design to **sustain, heal, and empower** our communities, as well as to seek liberation from exploitative and oppressive systems.

Principle 2

We **center the voices of those who are directly impacted** by the outcomes of the design process.

Principle 3

We **prioritize design's impact on the community** over the intentions of the designer.

Sean Dorr. "Making More Inclusive Forecasts"

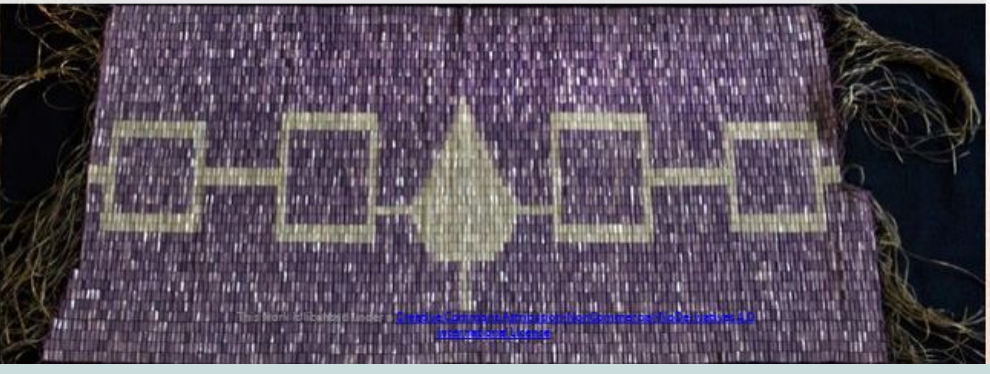
(there are 10 principles)

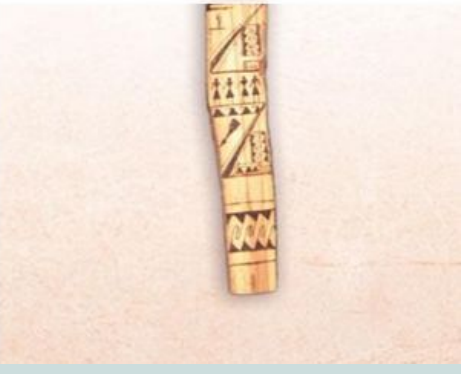
designjustice.mitpress.mit.edu

indigenous data sovereignty

Slides from: Global Indigenous Data Alliance. (2022). ‘Indigenous Data Sovereignty and Governance.’

OUR PEOPLES HAVE ALWAYS BEEN DATA EXPERTS





https://www.gida-global.org/care

Findable	Data has rich metadata and unique identifiers
Accessible	Data can be easily downloaded or used by using standard protocols
Interoperable	Metadata use an accessible and standard language
Reusable	Data is well-described and provides clear usage of licenses

https://www.gida-global.org/care

WHAT ARE **INDIGENOUS DATA?**

Data, information and knowledges, in any format, that impacts Indigenous Peoples, nations, and communities at the collective and individual levels:

USINDIGENOUSDATA.ORG
@USIDSN

Informed by British Columbia First Nations Data Governance Institute - BCFNDGI.COM

GIDA-GLOBAL.ORG
@GidaGlobal

https://www.gida-global.org/care

Collective Benefit	Data enables Indigenous peoples to derive benefit
Authority to control	Indigenous peoples rights and interests must be recognized - they determine how they are represented and identified in data
Responsibility	Those working with Indigenous data are responsible to share how it is used
Ethics	Indigenous People's rights and wellbeing should be the primary concern at all stages of data life cycle

https://www.gida-global.org/care

example of Indigenous data sovereignty: data about food

What is indigenous food sovereignty?

As stated in the Declaration of Nyéléni, there are six proposed principles of food sovereignty: focusing on food for people, valuing food providers, localizing food systems, making decisions locally, building knowledge and skill, and working with nature.

According to Dr. Cutcha Risling Baldy and Dr. Kaitlin Reed, the Co-Directors of the Food Sovereignty Lab at Cal Poly Humboldt, food sovereignty is also about centralizing Indigenous voices in how we collectively move forward in building sustainable food systems.


WesternLaw.org


American Rivers
OREGON
CALIFORNIA
Pacific Ocean
Williamson
Wood
Sycan
Chiloquin
Upper Klamath Lake
Sprague
Klamath Falls
Altamont
Lost
Keno
Willow Cr.
Boles Cr.
Tule Lake
Clear Lake
Lost Cr.
Butte Cr.
Shasta
Happy Camp
Yreka
Montague
Fort Jones
Klamath
Weed
Scott
Salmon
New
North Fork
Willow Creek
Trinity Lake
Weaverville
Trinity
South Fork
Hayfork Cr.
Hayfork
About 150 miles (241 km) across


Lost Coast Outpost


Times Standard

How does food sovereignty show up in your community?


Office of Minnesota Secretary of State


Animalia Life


MPR News

example of data and food sovereignty: mercury pollution

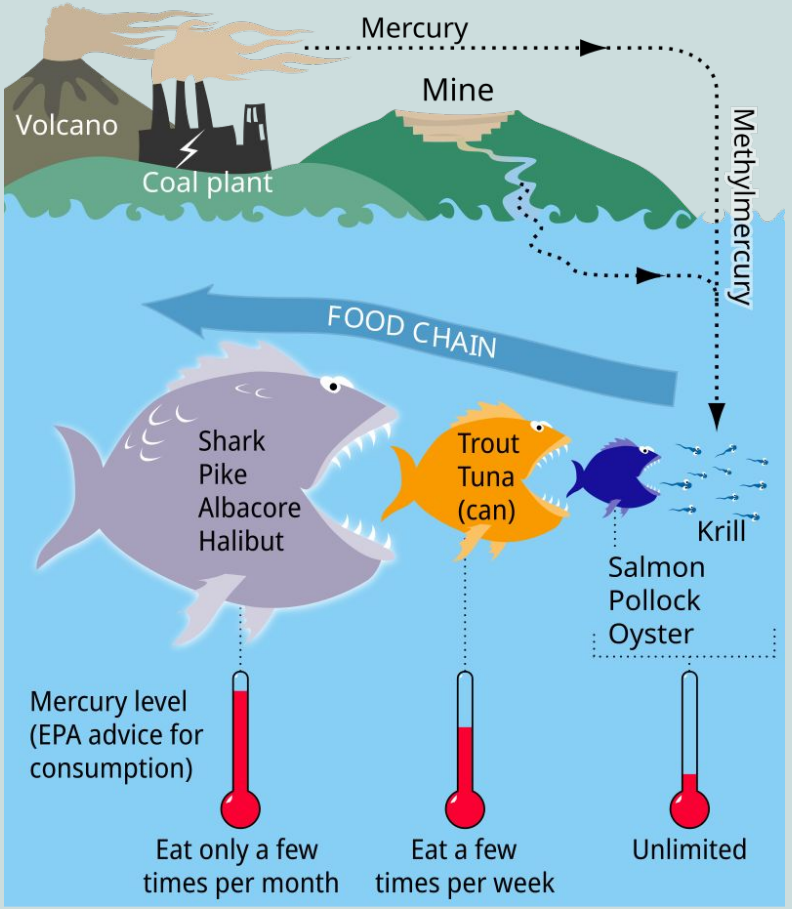


image source: wikipedia

Mercury pollution impacts fishing communities all over the world

Phys.org + Follow 38.1K Followers

Global mercury levels in rivers have doubled since Industrial Revolution, research reveals

Story by Science X staff • 1w • 3 min read

MONGABAY Features Videos Podcasts Specials Articles Shorts Donate

Mercury from gold mining contaminates Amazon communities' staple fish

Study Shows Mercury Levels in Arctic Wildlife Could Rise for Centuries

Mercury pollution is worsening a mental health crisis in this Indigenous community

"Our way of life has been totally destroyed."

By ANITA HOFSCHNEIDER PUBLISHED AUGUST 4, 2023 1:20PM (EDT)

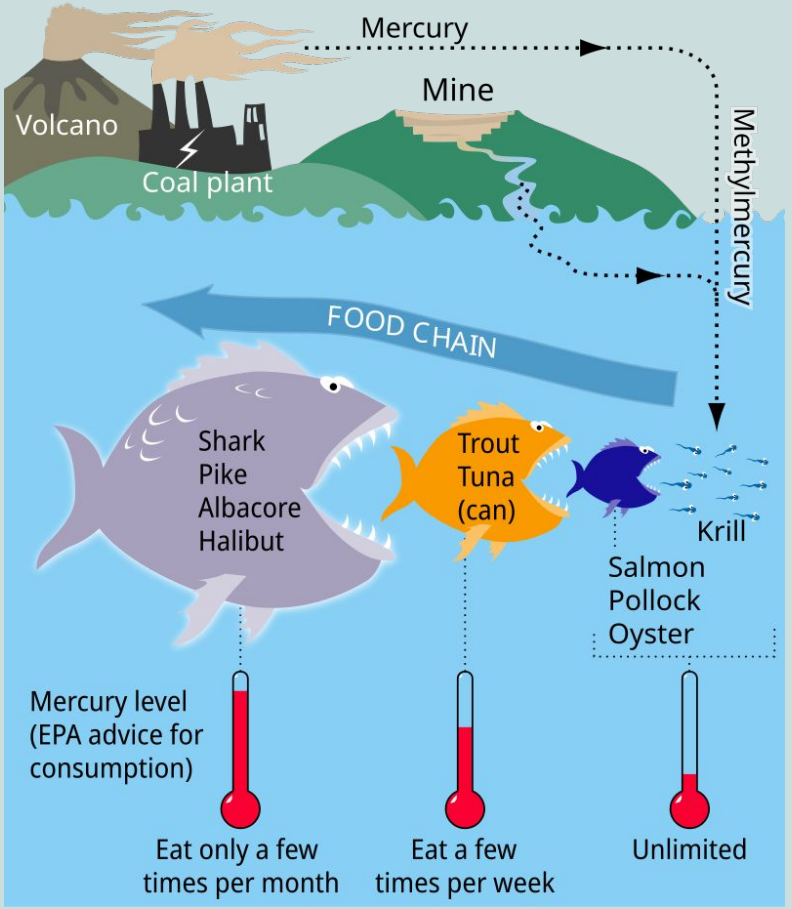
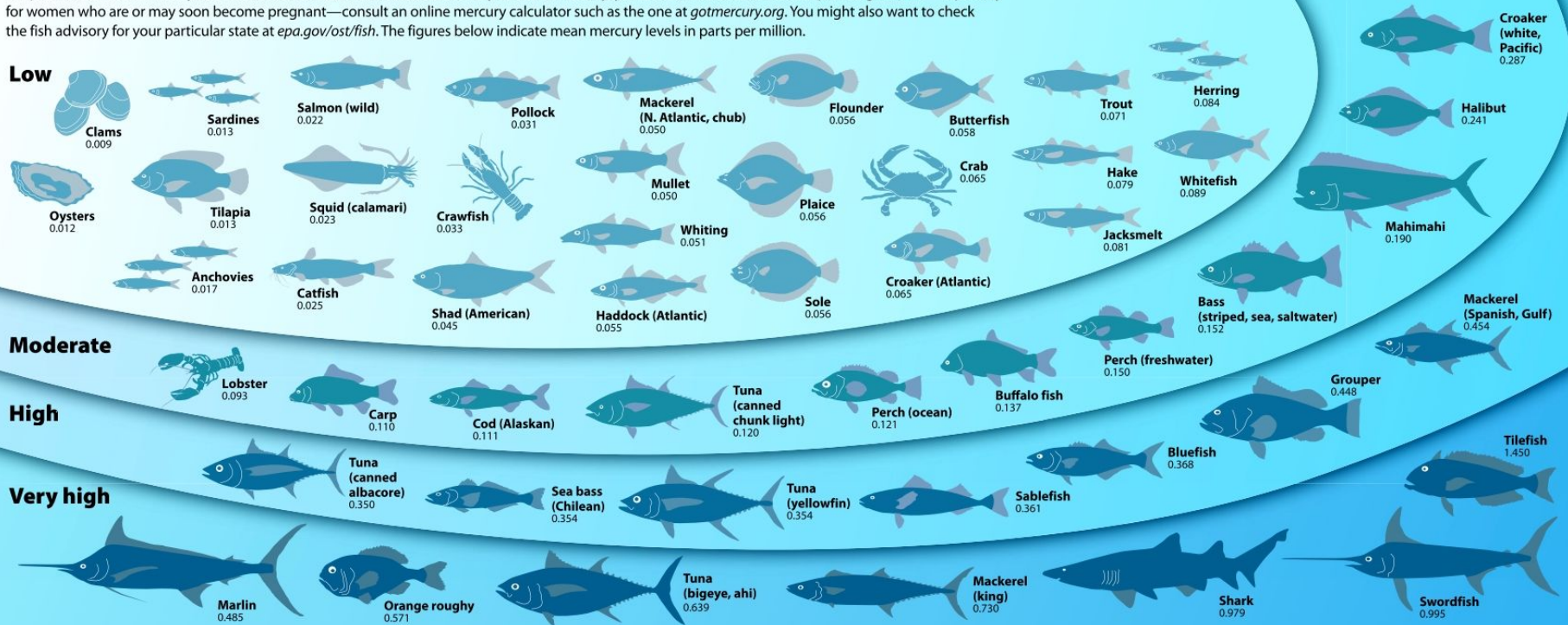


image source: wikipedia

But What Fish Can I Eat?

Despair not, fish lovers! Plenty of varieties of seafood are safe to eat, and many others can be enjoyed from time to time. For more precise guidance—especially for women who are or may soon become pregnant—consult an online mercury calculator such as the one at *gotmercury.org*. You might also want to check the fish advisory for your particular state at *epa.gov/ost/fish*. The figures below indicate mean mercury levels in parts per million.



How can we calculate mercury screening values for our own diets and our community's?

How are screening values calculated? FDA recommendations

Screening Values (SV) depend on:

- Reference dose (RfD) set by EPA
- body weight (BW)
- consumption rate (CR)
  - size of serving
  - servings per week

*Daily consumption rate (g/d) = serving size (ounce (oz)/serving) * (28.3 g/oz) * weekly servings (servings/wk) * (1wk/7d)*

$$SV = (RfD * BW) / CR$$

Table 1. Screening Values for Fish Categories

Weekly fish servings	Screening value (µg/g)	Chart category
0	> 0.46	Choices to Avoid
1	≤ 0.46	Good Choices
2	≤ 0.23	Good Choices
3	≤ 0.15	Best Choices

We can use R programming to calculate mercury screening values and explore data

What is R?

- R is a free software environment for statistics and plotting.
- Used by many ecologists, environmental scientists, statisticians
- Rstudio is a user interface that makes working with R easier

Goals for this morning

1. Identify the function name that loads packages
2. Recognize objects including variables, vectors, and data frames
3. Calculate values using mathematical operations and functions
4. Generate a scatterplot using ggplot2
5. Describe the relationships between body weight, weekly servings, and screening values (safe limits)

Let's get setup in R

Sign into an online account with posit cloud: https://posit.cloud/

Go to: hum.link/slawr-reu-sdi

credit: Allison Horst (https://allisonhorst.com/everything-else)

Getting Started–download R and R Studio

Step 1: Install R

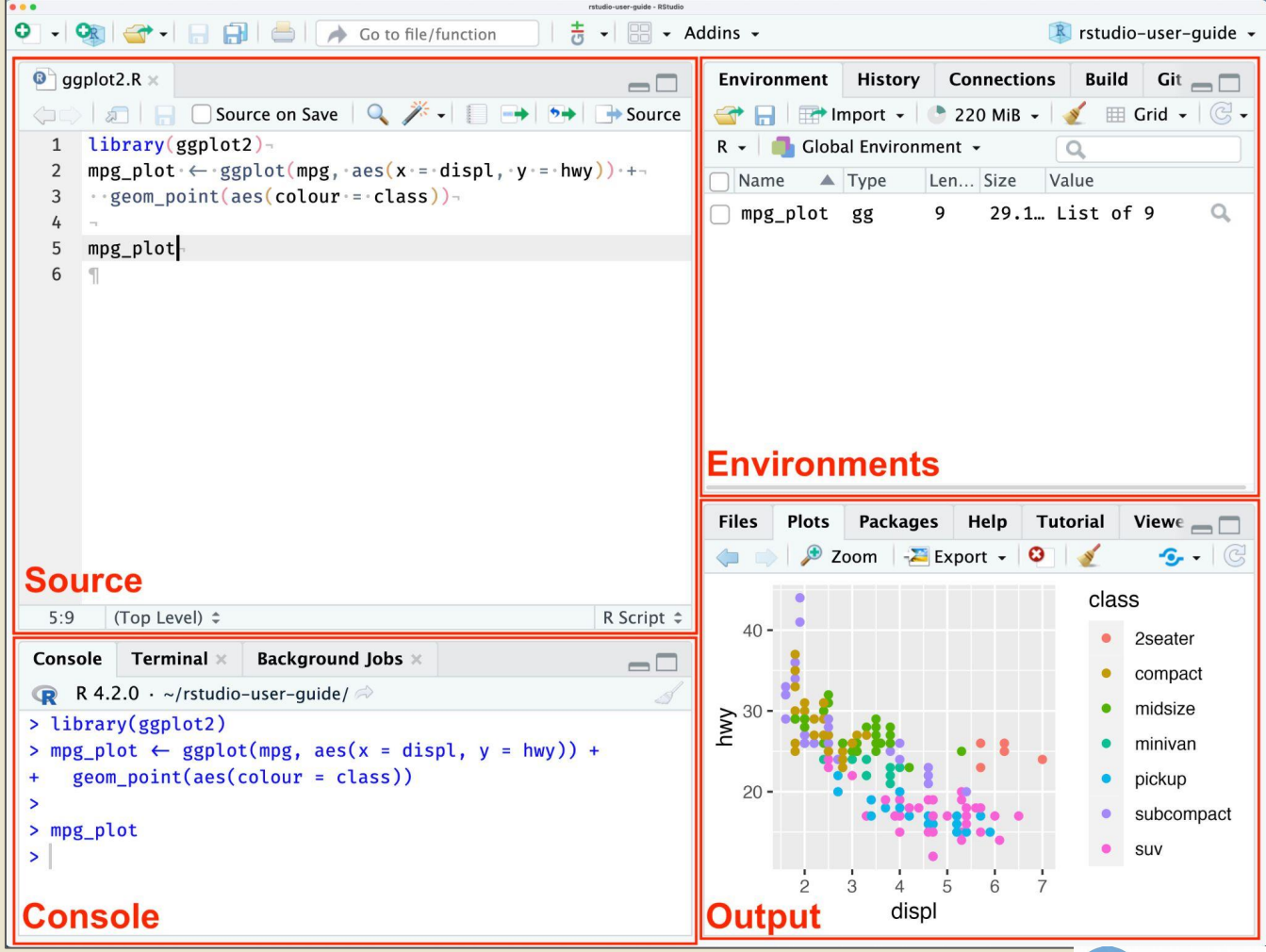
1. Go to the official R download page:
   👉 https://cran.r-project.org/
2. Click **"Download R for Windows"**.
3. Click **"base"**, then download the latest **R-x.x.x-win.exe** file.
4. Run the installer and follow these steps:
   - Click **Next** until you reach the installation location.
   - Use the **default settings** unless you have a reason to change them.
   - Finish the installation.

Step 2: Install RStudio

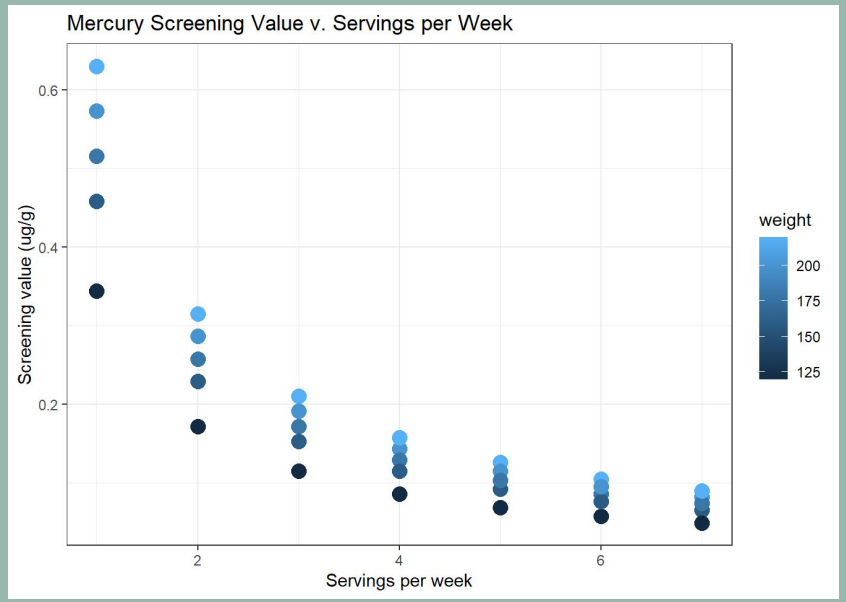
1. Go to the RStudio download page:
   👉
   https://posit.co/download/rstudio-desktop/
2. Download **RStudio Desktop (Free)** for Windows.
3. Run the installer and follow the instructions (keep default settings).
4. Once installed, open **RStudio**. It should automatically detect R.

Identify the 4 windows in Rstudio

1. Source Editor
2. Environments
3. Console
4. Output

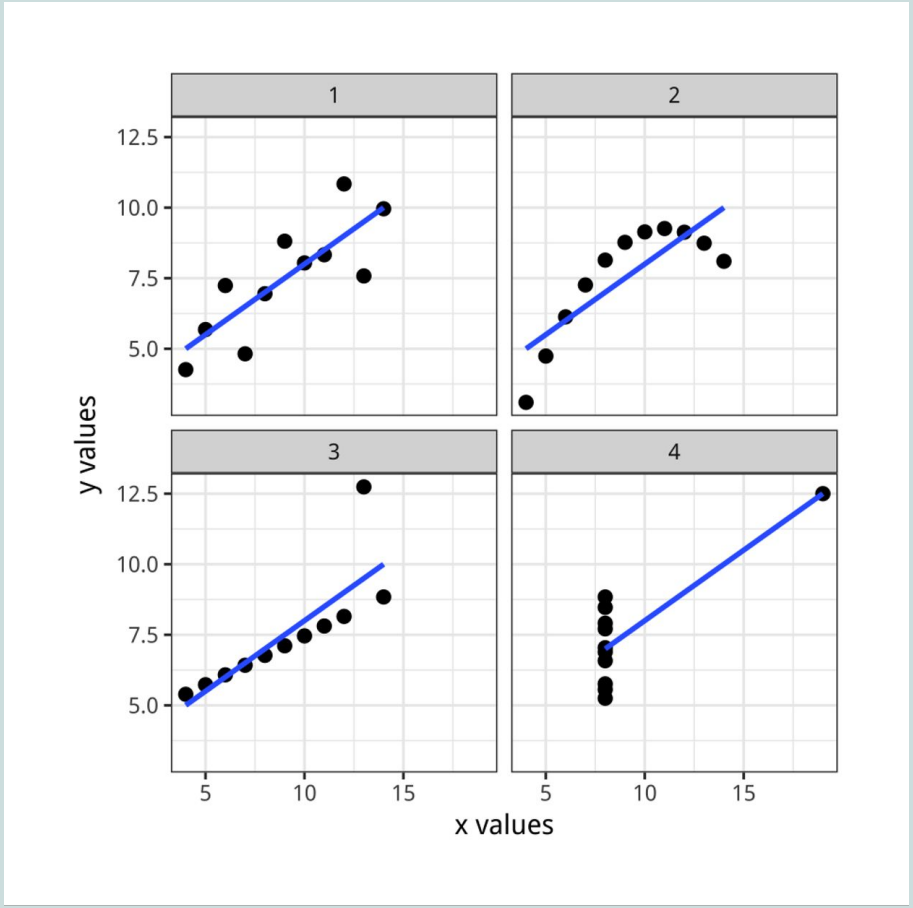


intro to data visualization



What are the goals of data visualization?

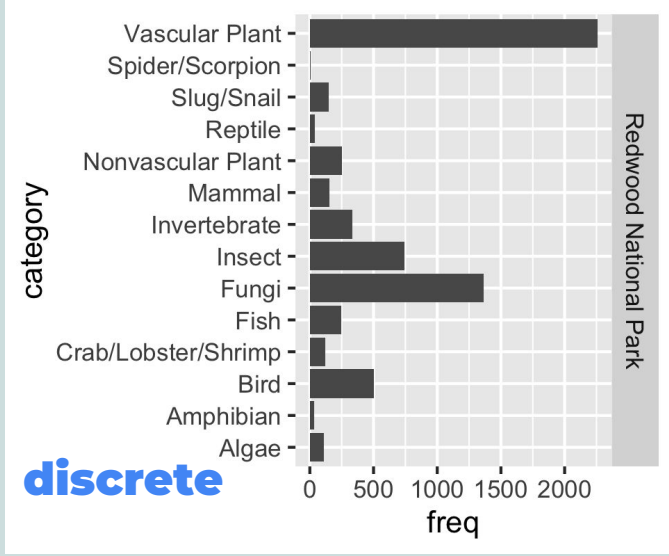
- exploring your data, finding patterns and relationships
- communicating your results, telling a story

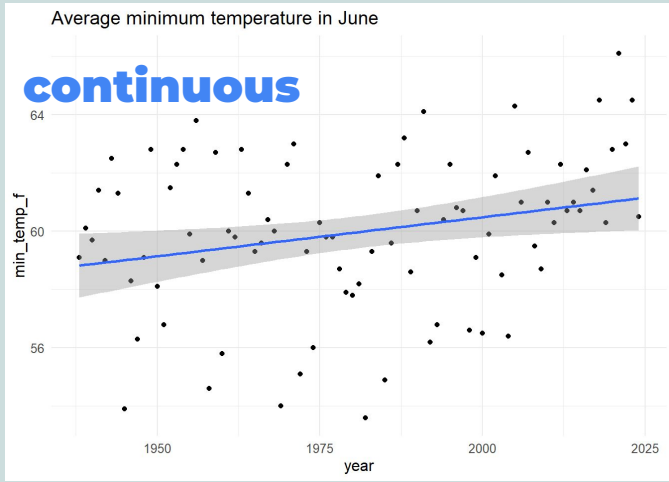


ANSCOMBE'S QUARTET (ANSCOMBE, 1973; CHATTERJEE & FIRAT, 2007)

Before starting a plot, know your data

- How many variables are there?
  - **1 or 2, more?**
- Discrete or continuous?
  - **discrete: land cover type, count**
  - **continuous: temperature, precip.**
- Are there categories or groups?
  - **for example: species, states**
- Is it temporal or spatial?





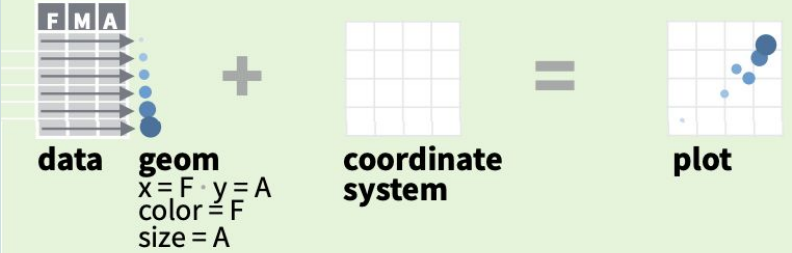
examples from Fundamental of Data Visualization by Claus O. Wilke

Before starting a plot, know your data

- How many variables are there?
- Discrete or continuous?
- Are there categories or groups?
- Is it temporal or spatial?

ggplot2 is based on the **grammar of graphics**, the idea that you can build every graph from the same components: a **data** set, a **coordinate system**, and geoms—visual marks that represent data points.

To display values, map variables in the data to visual properties of the geom (**aesthetics**) like **size**, **color**, and **x** and **y** locations.

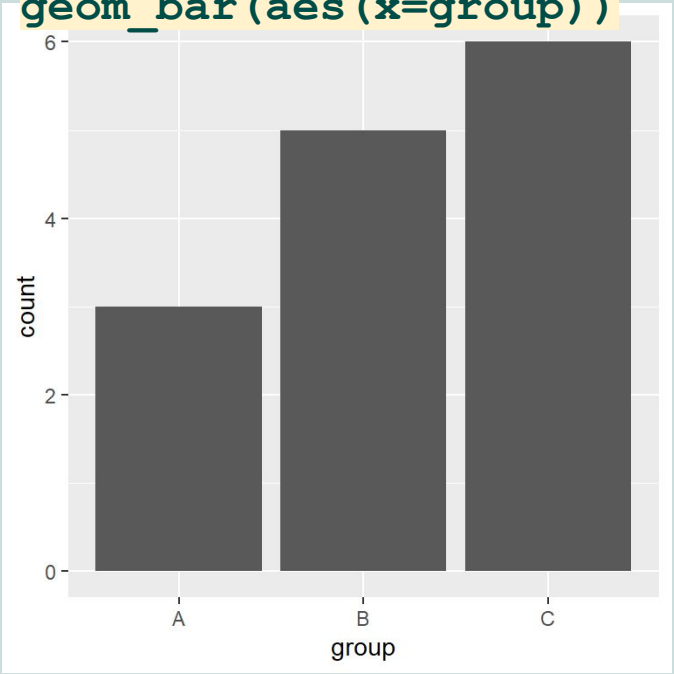


examples from Fundamental of Data Visualization by Claus O. Wilke

1 variable

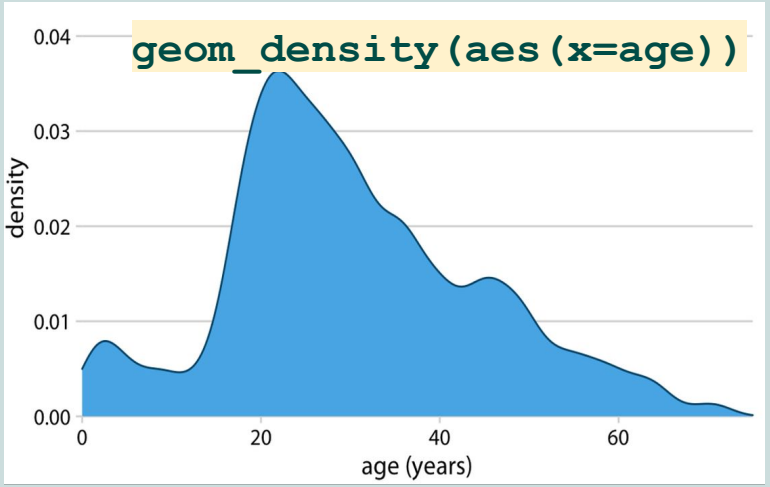
discrete - plot the count of occurrences

`geom_bar(aes(x=group))`

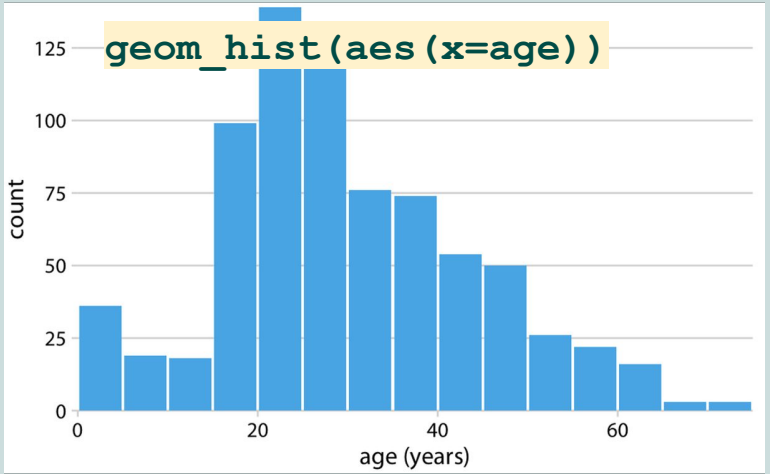


continuous - plot the frequency of values

`geom_density(aes(x=age))`



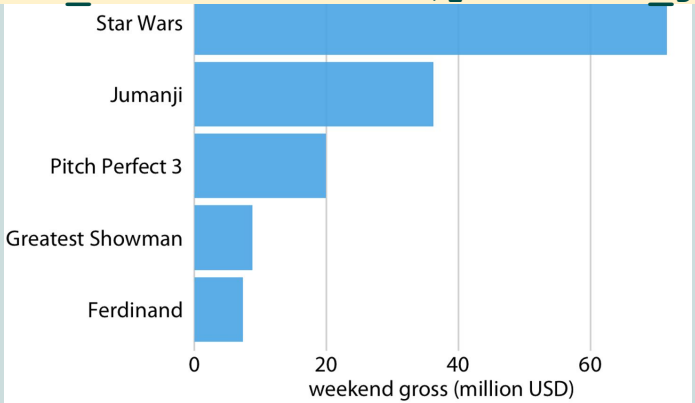
`geom_hist(aes(x=age))`

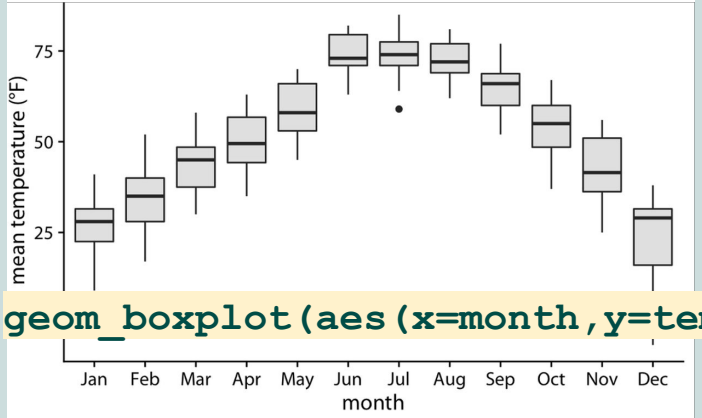


2 variables

discrete x, continuous y

```
geom_bar(aes(x=movie,y=weekend_gross))
```

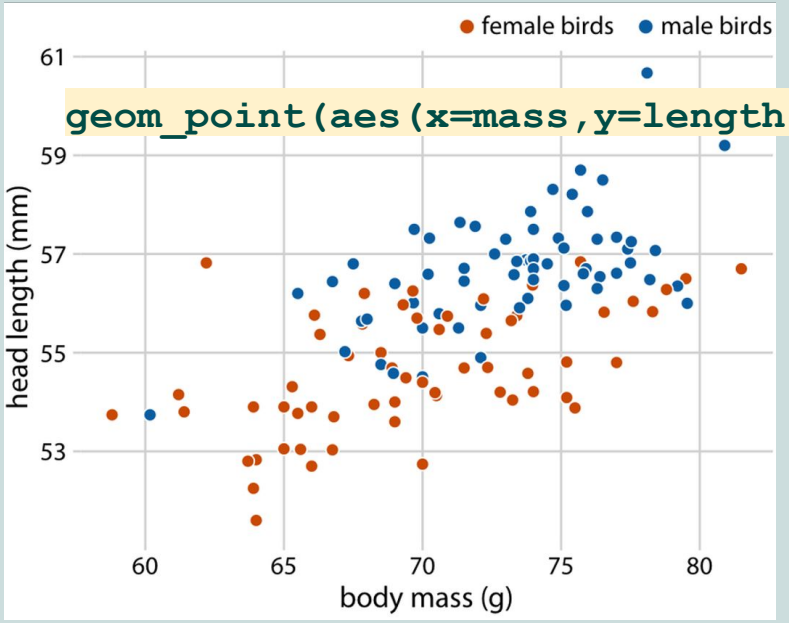





```
geom_boxplot(aes(x=month,y=temp))
```

continuous x, continuous y

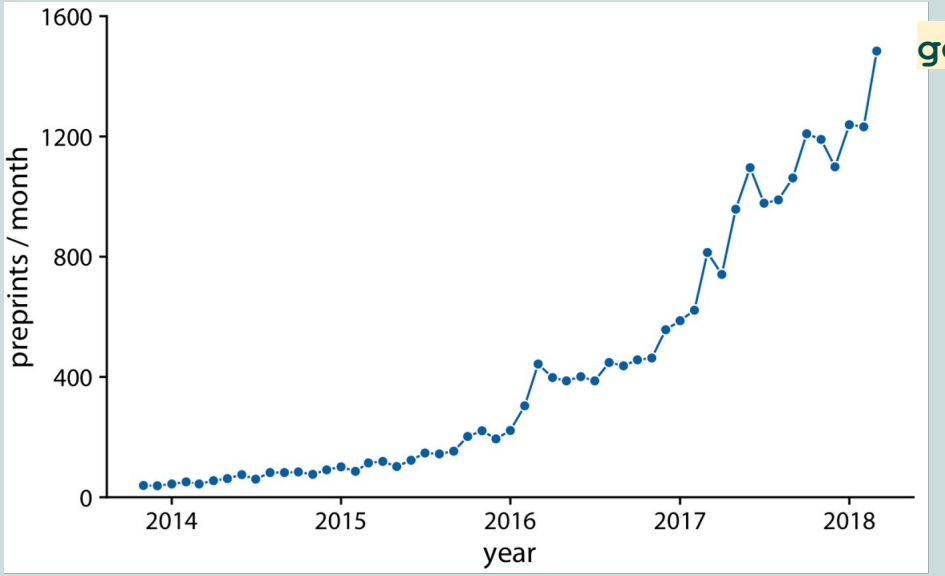
```
geom_point(aes(x=mass,y=length))
```



can also represent groups as color

time series

discrete ordered x, continuous y

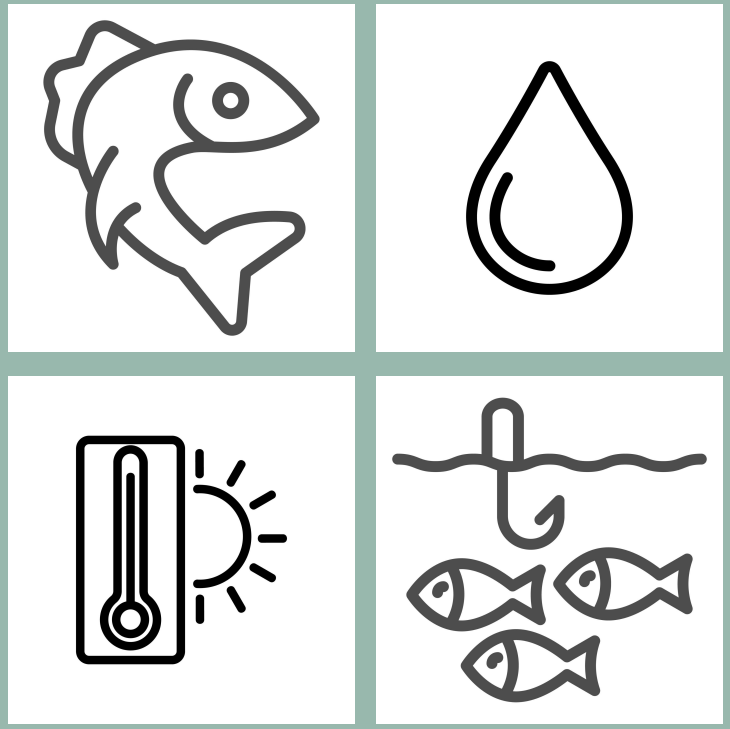



```
ggplot(data,aes(x=year,y=preprints))+
    geom_point() +
    geom_scatter()
```

What else should we consider?

What other tools for data visualization would be useful that may not be included using a programming language like R?

data vis case studies



data vis case studies

Open in R:

PM_data_vis_case_studies.Rmd

add your slides here:
hum.link/data-vis-slides

please send us feedback!

http://bit.ly/44iisdf