

Data Literacy

and Philippine Public Procurement Data

Ben Hur Pintor

Day 2 | 20-22 June 2022



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Outline

- I. About the event
 - A. Code of conduct
 - B. About me
 - C. About you
 - D. Objectives and expectations
 - E. Schedule
- II. Session 1
 - A. Gentle introduction to open data
 - B. Open contracting and the OCDS
- III. Session 2
 - A. What is data literacy?
 - B. How do we build individual and organizational data literacy?
- IV. Session 3
 - A. The Data Pipeline
 - B. Data-driven projects using the Data Pipeline
- V. Session 4
 - A. How to make better data presentations
 - B. Free and open source tools for working with data
- VI. Session 5
 - A. Data ethics, data fallacies, and other considerations
 - B. How to start sharing and opening data
- VII. Session 6
 - A. Data expedition and brainstorming a data-driven procurement-themed project



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

About me



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz

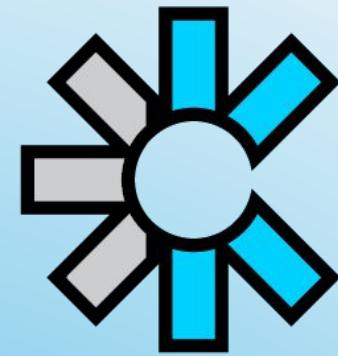


<https://bnhr.xyz>

Ben Hur Pintor

geospatial generalist. open stuff advocate. maptivist/dataactivist.

Data Training Lead



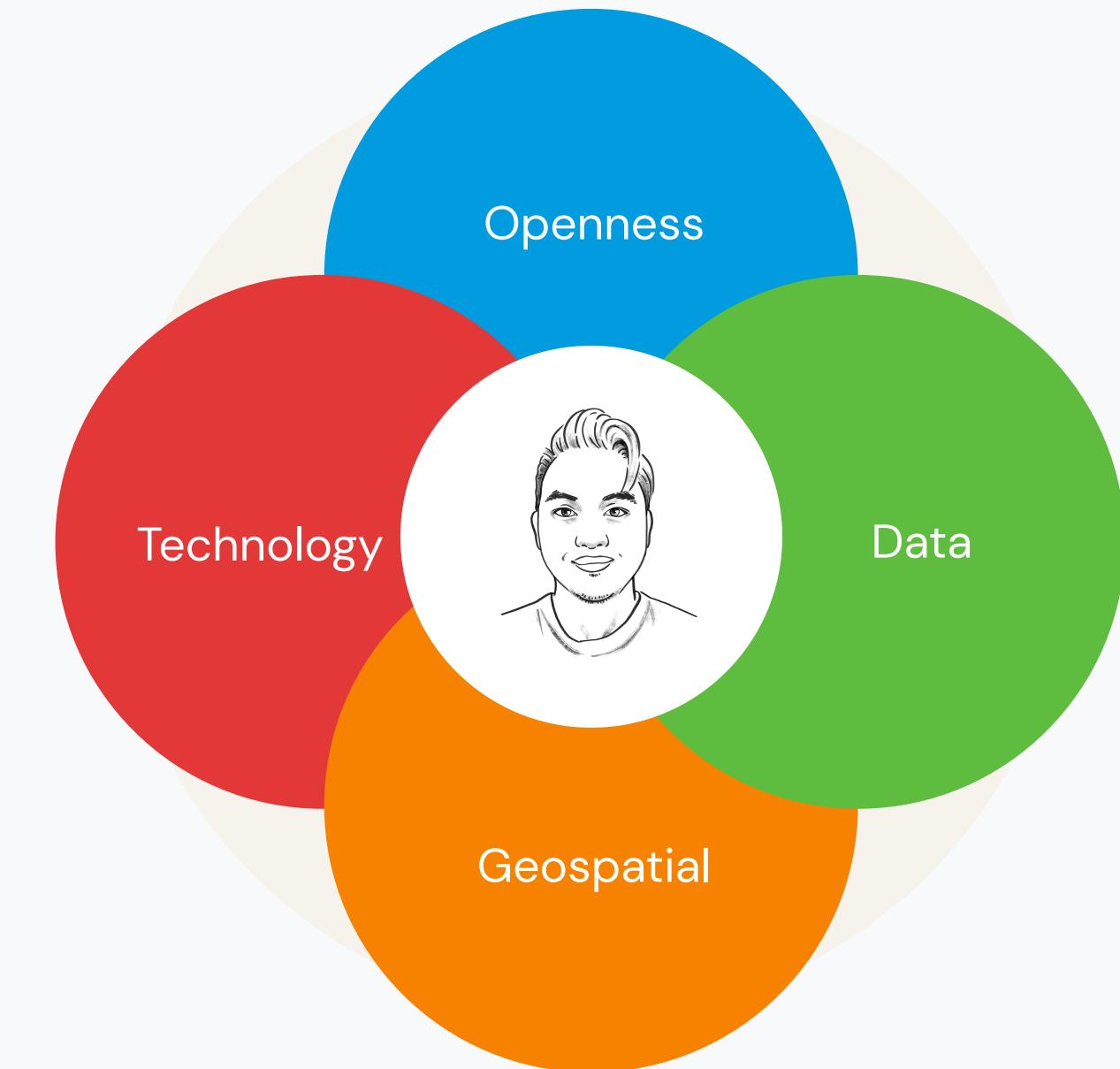
Open Knowledge
Foundation

Proprietor



openness.
data.
geospatial.

Chief Technology Officer



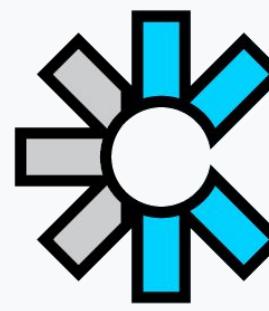
liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>



**Open Knowledge
Foundation**



We teach

If you or your organisation wants to learn about data literacy or even develop your skills to expert level, our team is here to help you on your journey.



We build

As open experts, we can create tools and provide services that help people and organisations put their data literacy learnings to work.



We organise

Through campaigning and community building, we're making an open future.



About SmartCT

SmartCT is the first tech non-profit in the Philippines and a pioneer in the smart cities field.

We aim to create a **movement that transforms the way we think, do, and plan smart cities and communities** especially in developing countries such as the Philippines through a co-developed, citizen-centric approach that puts openness and citizens at the heart of the development.

Making Smart Cities Open.



openness.
data.
geospatial.

- Established in 2019
- Part enterprise, part advocacy
- Provides training, support, and consulting services on open data, open source, data literacy, and free and open source software for geospatial applications (FOSS4G)

QGIS Certifying Organization

- Courses are vetted by the QGIS Project Steering Committee
- Can issue official QGIS Certificate

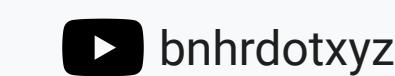
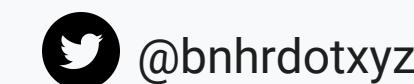
QGIS Sustaining Member

- Financially supports the QGIS Project
- Currently the only (and probably first) sustaining member from the Philippines

Some people I've worked with:



liberty. data. geospatial.



Recap of Day 1



liberty. data. geospatial.



@bnhr.xyz



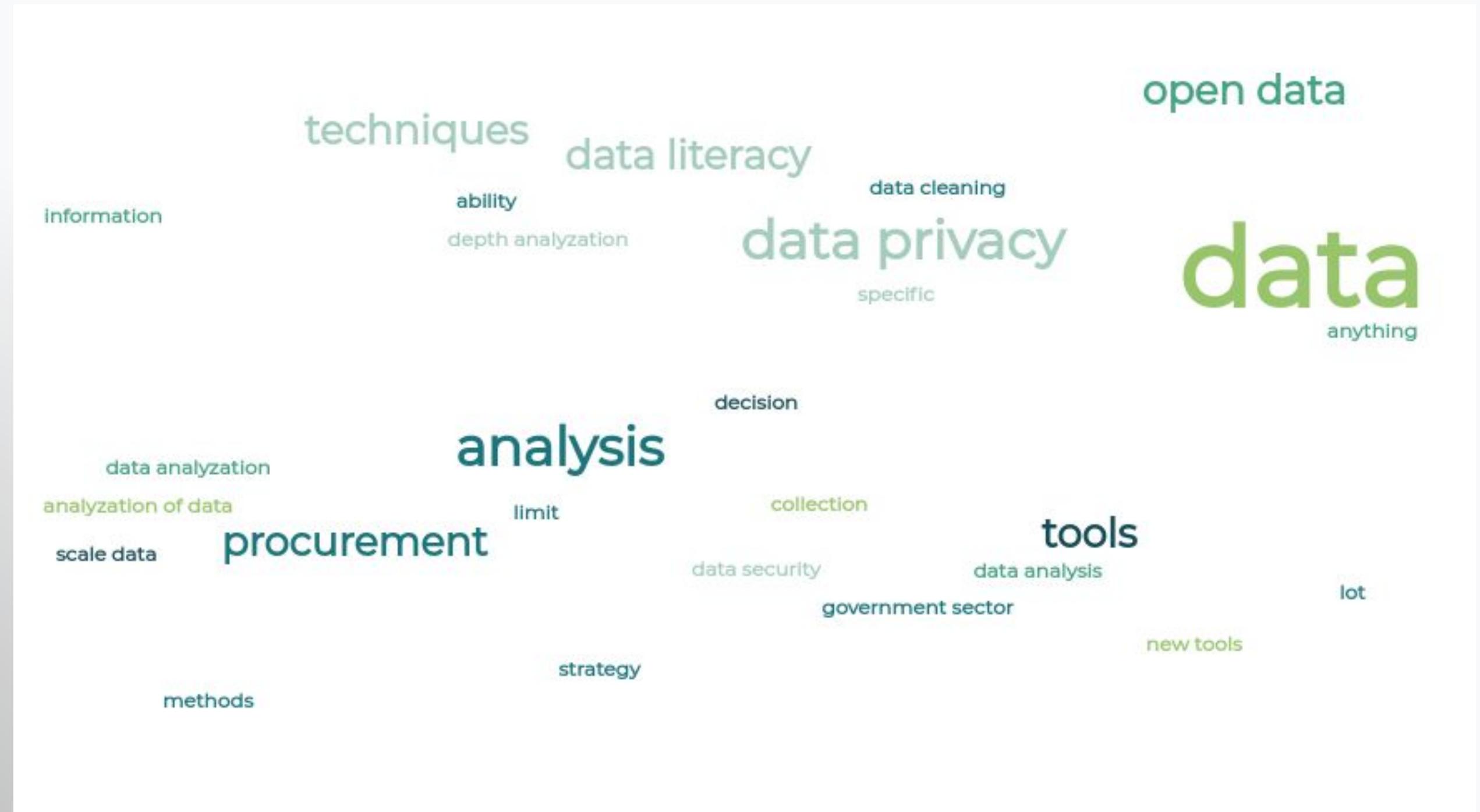
@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Any questions?



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Today's learning objectives

1. Be familiar with the steps of the Data Pipeline.
2. Learn the important aspects—best practices, tools, concepts, common mistakes, etc.—of each step of the Data Pipeline especially in the context of a data-driven project.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

The Data Pipeline

and doing data-driven projects



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

**But I'm not a data scientist /
I'm scared to work with data.**



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

What is the Data Pipeline?

The **Data Pipeline** is an approach developed by the School of Data network to **work with data from beginning to end**.

- Simple enough for beginners to grasp.
- Flexible enough for experienced practitioners innovate and adapt to different contexts.
- Utilized, extended, and improved upon by countless data practitioners over the years.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Steps in the Data Pipeline

The Data Pipeline divides the process of working with data into easily manageable steps:

- Define
- Find
- Get
- Verify
- Clean
- Analyze
- Present



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

The steps are not always linear

- You may need to do the find and get steps multiple times or repeat the verify and clean steps if you find mistakes in the outcome of your analysis.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Using the Data Pipeline

- **tool for teaching** how to work with data
- **guide for doing data-driven projects**



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Define



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Define

1. Start with a theme.
2. Break it down into specific questions that can be answered by data.
3. The questions should give you an idea of the kind of data or analyses you need.



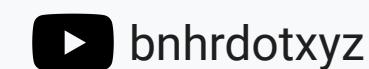
liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Define | example

Theme: COVID-19's impact on procurement practices in the Philippines

1. Are there significant differences between the prices of the commonly procured items during the pandemic and their prices pre-pandemic?
 - a. *What are the most commonly procured items during COVID-19?*
 - b. *At what price are these items currently being procured?*
 - c. *At what price were they being procured before COVID-19?*
 - d. *Are there differences in the prices and are they significant?*



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

DO: Formulate the research question(s) properly

- Data-driven projects are, by nature, investigative and seek to answer questions.
- By starting with a question or questions, you mitigate the tendency to design the project based on a predefined answer or solution
- **Visualization project:** research questions will help guide your thinking about what should be visualized
- **Advocacy campaign:** research question will help focus on the message that the campaign aims to highlight



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

DO: Identify the scope and challenges

- A properly formulated research question will highlight the scope of the project and the data needed.
- Are there going to be challenges in getting the data?
 - unresponsive or uncooperative government agency
 - data collection in a remote location



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

DON'T: Proceed without a research question

- A project can have several research questions but having none shows a lack of rigor or direction.
- **Typical consequences include:**
 - Dashboards designed around the data instead of the target audience.
 - Delayed projects due to bad scope management.
 - Difficulty to pivot the project if some of the expected data is not available.



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

AVOID: Bad scoping

- Don't forget to account for other variables linked to the sources of data, its quality, and its accessibility.
- **Typical consequences include:**
 - Delays in the project delivery due to lack of anticipation of the difficulty of getting access to data.
 - Lower quality of project outputs due to overestimation of the quality of the data
 - Delays or low quality data due to underestimating the time needed to clean a dataset manually



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

Any questions?



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Find



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Find

1. Know where to find data and how accessible it is.
2. Be creative and critical. When data seems hard to find, you can consider looking at proxy indicators—an indirect measure or sign that indicates a phenomenon in the absence of a direct measure or sign.



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

DO: Ask the right questions

- Several sources may maintain similar datasets
- Understand the precise data needs of your project
- **When looking for data, you can look at:**
 - national government agency mandated to collect the data
 - organizations or individuals who are working or have worked with the data
 - other datasets that can serve as indicators for
 - online search engines (utilizing advanced search capabilities makes searching easier)
 - libraries and museums.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

ASK: Where can you find the data?

- Data portals
- Digital systems
- Physical archives
- Behind an API
- On the internet
- Nowhere



Aside from the location of the data, it is also important to note the data format of the data you've found—it will serve you well in the next step.



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

KNOW: Data formats

- It is important that you know about different file formats so that you can better plan a strategy on how to get the data
- Common file formats/extensions include:
 - .txt
 - .csv/.tsv
 - .xls/.xlsx
 - .ods
 - .json
 - .pdf



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

Activity: Using search operators

Search operators allow you to refine your search so you can easily find what you are looking for. It follows the format <operator>:<value> and is included in the text you use for your search. Example operators include:

- site - specify a URL domain or subdomain (.gov.ph)
- filetype - specify the filetype (e.g. pdf, xlsx)
- "<text>" - search for an exact match of <text>

Exercise:

1. Go to your favorite search engine.
2. Search for the following:
 - a. site:.gov.ph filetype:xlsx "philippine population"
 - b. site:.gov.ph filetype:xlsx "procurement"
 - c. site:.gov.ph filetype:pdf "procurement"
3. Try other searches with other search operators



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

DON'T: Settle for a poor quality dataset

- you will find multiple datasets and sources
- avoid settling for the first dataset you find
- may be more useful to create the needed dataset out of several quality datasets



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

Any questions?



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Get



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Get

1. How do you get the data from its source/location into my computer/machine?
2. Do you need convert it into a different format to make it machine-readable?



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

KNOW: Key terms and technologies

- **web scraping** - the automated retrieval information on a webpage
- **pdf extraction** - automatically extracting data tables from PDF
- **OCR (optical character recognition)** - allows the extraction of texts from images or scanned documents
- **API (application programming interface)** - a way for people to query and request for information from a dataset/database



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

KNOW: Methods of getting data

- **Get the data directly**
 - downloading from open data portals
 - using an Application Programming Interface (API)
- **Request for the data**
 - sending an FOI request or a formal request letter to the data owner
- **Extract data from files and documents**
 - extract tables from PDFs using applications such as Tabula or Excalibur
 - scrape tables from webpages using Google Sheets, webscraper.io, or programming with Python
 - extracting texts from scanned documents or images using OCR
- **Do the data collection yourself**
 - Manual pen-and-paper surveys
 - Mobile data collection applications such as ODK/ODK-X, KoBo, or Sabasi
 - Crowdsourcing via socmed or platforms such as mTurk and samasource



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Activity: Using an API

An API is the intermediary between you and a database you are requesting data from. You can think of it like this: bank = database, your money = data, withdrawal form = API.

Exercise:

1. Go to <https://genderize.io/>.
2. This API predicts the gender of a person given their name based on a dataset that contains names, genders, and other information.
3. Try the different API endpoints (or requests) available.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Activity: Scraping with GSHEET

Exercise:

1. Open Google Sheets.
2. Go to
<https://www.worldometers.info/world-population/philippines-population/>
 - a. We'll scrape the Population of the Philippines (2020 and historical) from this website into our spreadsheet using the **IMPORTHTML** function.
3. Use the following function in your spreadsheet:
=IMPORTHTML("https://www.worldometers.info/world-population/philippines-population/", "table", 2)
4. What happened?
5. What are the advantages of using this approach?



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

KNOW: Data collection

- when data does not exist or cannot be found
- <https://school-of-data.github.io/mobile-data-collection/index.html>



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

AVOID: Bad data collection methodology

There are two common types of data collection projects:

- The first one is similar to a **desk research** where the **goal is to create a dataset out of multiple data points from different sources**. This is **usually done online**.
- The other one involves **field data collection** and requires you to **send people with questionnaires** to fill either by themselves (e.g. doing rapid damage assessment after a disaster) or with the answers from a respondent.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

DON'T: No policy for data storage, access, and protection

- policies for the storage, access, and protection of the data
- data redundancy and backups
- who has access to what data
- how to keep sensitive data such as personally-identifiable information private and secure



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

Any questions?



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Verify



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Verify

1. How bad is the data?
 - a. Check internal consistency
 - b. Check for outliers
 - c. Consult metadata if available
2. How much can you trust the data?
 - a. Data trustworthiness, completeness, and quality
3. How can I verify the data?



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

KNOW: Data trustworthiness, completeness, and quality

- **Data trustworthiness** - how much we can trust the data to represent what it says it does
- **Data completeness** - how much the data covers the reality it tries to represent
- **Data quality** - how well the dataset is structured and documented



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

KNOW: Metadata

- data about the data
- includes every relevant contextual information about the dataset—who, what, when, where, why, how the data was created
- need not be complex
 - stored alongside the data (in different tab of a spreadsheet)
 - shared alongside the dataset (as a json, doc, or pdf)



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

KNOW: The four verification methods

- 1. Asking the source:** the ones who produced or published the data are most likely the best experts on it
- 2. Asking experts:** there is probably someone can help you better contextualise the dataset for you
- 3. Statistical checks:** finding outliers and weird values or patterns, describe the data (mean, min, max, etc.)
- 4. Common sense check:** relies on having a general sense for reading data but also good background knowledge about the context of the data



DON'T: Skip the verification process

- ideally, every new dataset must go through at least one of the four verification methods before it is used
- lack of verification can lead to missed inconsistencies and issues in the data
- [The Quartz guide to bad data](#)



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

Any questions?



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Clean



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Clean

All data is dirty but just because it is dirty does not mean it's garbage.

- majority of the work
- time-consuming and can also be error prone
- works as a group with the VERIFY and ANALYZE steps
 - cleaning the dataset a bit to be able to verify its content
 - going back to the verify step after finding something strange during the analysis
 - doing some basic analysis steps as part of the verification process



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

KNOW: Goals of data cleaning

- create a consistent, human understandable, machine readable dataset
- prepare the data for a specific analysis or visualization you want to use the data for



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

KNOW: Data cleaning steps

1. **Data tidying:** cleaning the structure of the data without touching its content and following the principles of tidy data.
2. **Data editing:** modifying the content of the data to prepare it for your analysis. This involves correcting problems with the values stored in the dataset.
3. **Data consolidation:** adding complementary data to your main dataset. This step provides an opportunity to complement or extend the dataset that you have collected, verified and cleaned.



KNOW: Best practices

1. Always back-up your data.
2. Make a copy for every step of the cleaning process (e.g. a tab in a spreadsheet for each step, version control).
3. Avoid overwriting data.
4. DO NOT DELETE DATA!
5. Documentation!



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

KNOW: Tools and techniques

- **Spreadsheet applications:** LibreOffice Calc, Google Sheets, Microsoft Excel
- **Specialized tools:** OpenRefine
- **Programming:** Python, R, etc

Which tool should I use?

- Stick with what you're comfortable with
- Start with the simplest tool first



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

KNOW: Things to clean

1. **Formatting problems:** related to HOW the dataset is structured. These are usually resolved in the data tidying step.
2. **Content problems:** related to WHAT is written or stored in the dataset. These are usually resolved in the data editing step



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

Formatting: Data stored in multiple columns

Example:

Address is divided into BARANGAY, MUNICIPALITY, and PROVINCE columns but it needs to be a single string.

	A	B	C
1	Barangay	Municipality	Province
2	NEW ALABANG VILLAGE	CITY OF MUNTINLUPA	NCR FOURTH DISTRICT
3	GREENHILLS	CITY OF SAN JUAN	NCR SECOND DISTRICT
4	SALVACION	QUEZON CITY	NCR SECOND DISTRICT

Solution:

Concatenate cell values into a single cell.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Formatting: Data stored in multiple columns

Google Sheets

	A	B	C	D
1	Barangay	Municipality	Province	NEW ALABANG VILLAGE, CITY OF MUNTINLUPA, N...
2	NEW ALABANG VILLAGE	CITY OF MUNTINLUPA	NCR FOURTH DISTRICT	=CONCATENATE(A2, " ", B2, " ", C2)
3	GREENHILLS	CITY OF SAN JUAN	NCR SECOND DISTRICT	

LibreOffice Calc

	A	B	C	D
1	Barangay	Municipality	Province	Address
2	NEW ALABANG VILLAGE	CITY OF MUNTINLUPA	NCR FOURTH DISTRICT	=CONCATENATE(A2, " ", B2, " ", C2)
3	GREENHILLS	CITY OF SAN JUAN	NCR SECOND DISTRICT	
4				CONCATENATE(Text1, Text2, ▶ ...) : Text for the concatenation.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Formatting: Data stored in a single column

Example:

Address is stored as a single string of BARANGAY, MUNICIPALITY, PROVINCE but you need to summarize data based on BARANGAY only

	D	E	F
1	Address	Individual Population (as of 01Jul2007)	Individual Population (as of 01May2010)
2	NEW ALABANG VILLAGE, CITY OF MUNTINLUPA, NCR FOURTH DISTRICT	20226	20349
3	GREENHILLS, CITY OF SAN JUAN, NCR SECOND DISTRICT	16632	12548
4	SALVACION, QUEZON CITY, NCR SECOND DISTRICT	9211	8194



Solution:

Split the column into several columns



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Formatting: Data stored in a single column

For spreadsheets:

1. Add X columns after the column to be split
 - a. X = number of columns to split text to
2. Copy the column to be split into the first column
3. Select the cells to split
4. Run:
 - a. **Data -> Split text to columns** (Google Sheets)
 - b. **Data -> Text to columns** (LibreOffice Calc)



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

Content: Extra white spaces

Example:

	A		
1	Column		
2	tHiS hAs	ExTrA	sPaCeS
3	tHiS hAs	ExTrA	sPaCeS
4	tHiS hAs	ExTrA	sPaCeS
5	tHiS hAs	ExTrA	sPaCeS



Solution:

Trim/remove the extra white spaces

If needed, you can overwrite the cell using Paste Special.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Content: Extra white spaces

For spreadsheets

1. Run **TRIM(<cell to remove extra spaces>)**
2. Copy trimmed text (CTRL + C)
3. Paste Special -> Paste values only (CTRL + SHIFT + V)

	A	B	C	D	E
1	Column				
2	tHiS hAs	ExTrA	sPaCeS		
3	tHiS hAs	ExTrA	sPaCeS		
4	tHiS hAs	ExTrA	sPaCeS		
5	tHiS hAs	ExTrA	sPaCeS		
6					
7					
8					
9					
10					



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Content: Wrong case

Example:

	A
1	Column
2	tHiS hAs ExTrA sPaCeS
3	tHiS hAs ExTrA sPaCeS
4	tHiS hAs ExTrA sPaCeS
5	tHiS hAs ExTrA sPaCeS



Solution:

Convert the text to the case of choice (lowercase, UPPERCASE, Proper Case)

If needed, you can overwrite the cell using Paste Special.

Content: Wrong case

For spreadsheets

1. Run **LOWER(<lowecase>)**, **UPPER(<UPPERCASE>)**, or **PROPER(<Proper Case>)**
2. Copy text
3. Paste Special -> Paste values only

C9	A	B	C	D	E	F
1	Column					
2	tHiS hAs ExTrA sPaCeS					
3	tHiS hAs ExTrA sPaCeS					
4	tHiS hAs ExTrA sPaCeS					
5	tHiS hAs ExTrA sPaCeS					
6						
7						
8						



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

ALWAYS: Keep track of changes

1. For **spreadsheets**, create a new tab every time an important modification is applied to the data.
2. More tech-savvy data practitioners might use a **version control system** like GitHub or GitLab for the same purpose.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Activity - Cleaning data

1. Go to this link <https://bit.ly/3N1S1hz> and make a copy of the spreadsheet: File -> Make a copy
2. Follow the instructions in the **INSTRUCTIONS** tab.
3. Good luck!



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

Any questions?



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Analyze



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



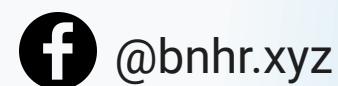
<https://bnhr.xyz>

Analyze

- Interview the dataset.
- Convert your questions to specific spreadsheet, math or statistics formulas.
- Find elements that answer your questions.
 - Minimum/Maximum
 - Least/Most
 - Order/Rank



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

KNOW: Type of analysis

- **Descriptive analysis:** describe the basic features or characteristics of the data, such as the mean, the median, the maximum, the minimum, etc.
- **Inferential analysis:** make reasonable guesses about your data and how it relates to the larger dataset that it is part of or other datasets that are similar to itself.
- **Predictive analysis:** make reasonable predictions about the probability of future or otherwise unknown events based on current or past data; advanced.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

KNOW: Tools and techniques

- The kinds of tools and techniques you will use depends on your goals and data. If you are working mostly with **tabular data in order to perform descriptive analysis**, a **spreadsheet application would suffice**.
- For a complex data project, a single tool may not be enough and you would need to use different kinds of tools and techniques to analyze your data.
- It's okay to seek help with tools that you may not be familiar with or are not comfortable using.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

KNOW: Tools and techniques

For more complex or specialized your goals and data are, you may need:

- **Programming languages** such as Python and R if you need complex, advanced, and automated computations.
- **Databases and SQL** if you are dealing with large amounts of data and you need a scalable and robust storage + analysis solution.
- **GIS and geospatial applications** if you are dealing with geospatial/location data, location is an important variable in your analysis, or you just want to make maps
- **Statistical software** such as SPSS and Stata for more advanced statistical analysis.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

DO: Use a hypothesis

- helps you interpret the results
- makes you more deliberate with your analysis
- prevents you from becoming arbitrary in your approach
- even if the results of your analysis tells you to reject the hypothesis you made, this outcome is still interesting because it means that the data does not say what you expected it to
- can be based on previous analyses or expectations of what the data should show



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

DO: Make the analysis reproducible

- document the steps
- follow FAIR principles
- use Frictionless technologies



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

AVOID: Not using a hypothesis

Example: What does it mean if, when studying public procurement, you find that 50% of contracts are won by only 10% of the total number of suppliers who submitted bids? Is it a lot or too little? Good or bad?

Making a hypothesis (e.g. there is healthy competition when at least 25% of the suppliers win 50% of the contracts), based on past data or prevailing assumptions about the topic, helps frame the results of the analysis.



Activity - Using Pivot Tables

1. Go to this link <https://bit.ly/3HFMYIN> and make a copy of the spreadsheet: File -> Make a copy
2. Go to this link <https://forms.gle/coZVzJDdXWqpURpd7> and answer the questions based on the data from 1.
3. Try to utilize pivot tables to make your life easier.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Any questions?



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Present



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Activity - Data presentation gallery

1. Go to this link <https://bit.ly/3OaNyu3> and answer the form.

Let's be back after 15 minutes.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Present

The use of the term data presentation instead of data visualization in the Data Pipeline is deliberate. The goal when working with data is to effectively present and communicate what the data says—and that doesn't always require the use of a chart or graph.

- the key parameters of data presentation are your **audience** and your **message**



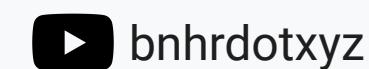
liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

KNOW: The goals of the project

- different goals = different presentation needs
- same dataset may be presented differently depending on the audience
 - **report:** you may choose to include a table and a simple graph
 - **specialized or expert audience:** you may prefer the use of more complex or interactive visualizations
 - **general public:** you might want to create an infographic



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

DO: Put the audience first

- think about your audience when planning or creating your data presentation
- consult and communicate with them at each step of your creative process
 - co-create
 - user-test
- the success of your data presentation is not up to you but to your audience



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

DO: Iterate

- don't rush to publish a visualization
- iteration allows you to refine your visualization
- the final data presentation you publish will be more robust and fit-for-purpose



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

KNOW: Choosing your visualization

- different visualizations have different strengths and weaknesses
- **how to choose?**
 - experience
 - studying and researching other visualizations
 - data visualization catalogs



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

KNOW: Tools and techniques

The tools and techniques you will use will depend on the presentation or visualization you want to create.

Spreadsheet applications

- [LibreOffice Calc](#)
- [ONLYOFFICE Spreadsheet](#)
- [Google Sheets](#)
- Microsoft Excel

GIS and mapping applications

- [QGIS](#)
- [Leaflet](#)
- [MapLibre](#)
- [Mapbox](#)

Web-based visualization and design applications

- [Datawrapper](#)
- [RAWGraphs](#)
- [Canva](#)
- [Google Data Studio](#)

Programming

- Python
- R

Design and image processing applications

- [GIMP](#)
- [Inkscape](#)
- [Krita](#)
- Adobe Creative Cloud products

Data visualization catalogs

- [The Data Visualisation Catalogue](#)
- [From Data to Viz](#)
- [Data Viz Project](#)
- [Chartmaker Directory](#)
- [Python Graph Gallery](#)
- [R Graph Gallery](#)



AVOID: Not knowing your audience

- visualisation is not meant to be a neutral presentation
- communicating to different audiences will require you to utilize different visualization techniques—a chart that works for one group of people may not work for another
- consider the data you have, the context of the presentation, and the data visualization literacy of your audience



liberty. data. geospatial.

@bnhr.xyz

@bnhrdotxyz

bnhrdotxyz

<https://bnhr.xyz>

Any questions?



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Making better data presentations



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Choosing your visualization



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

What kind of data do you have?

- Categorical
- Numeric
- Both categorical and numeric
- Time-series
- Geographic
- Network



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

What relationship will you emphasize?

- Correlation
- Ranking
- Comparison
- Distribution
- Map
- Part of a whole
- Evolution



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



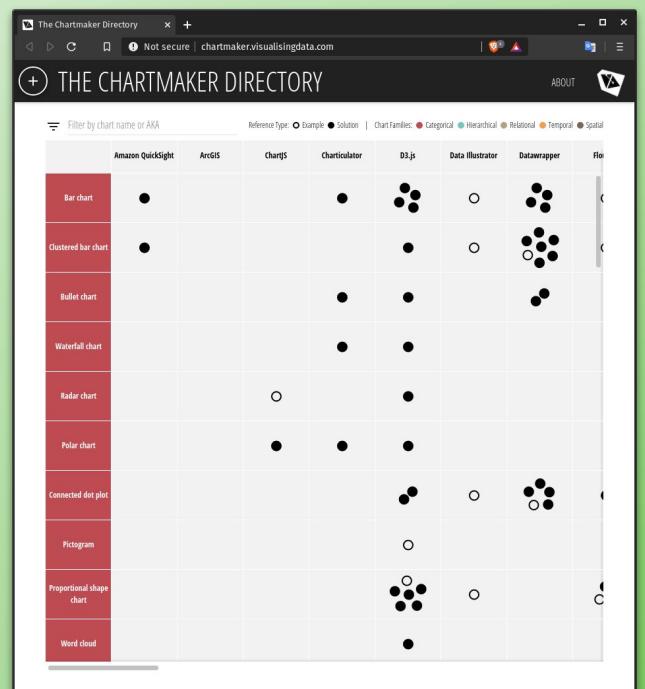
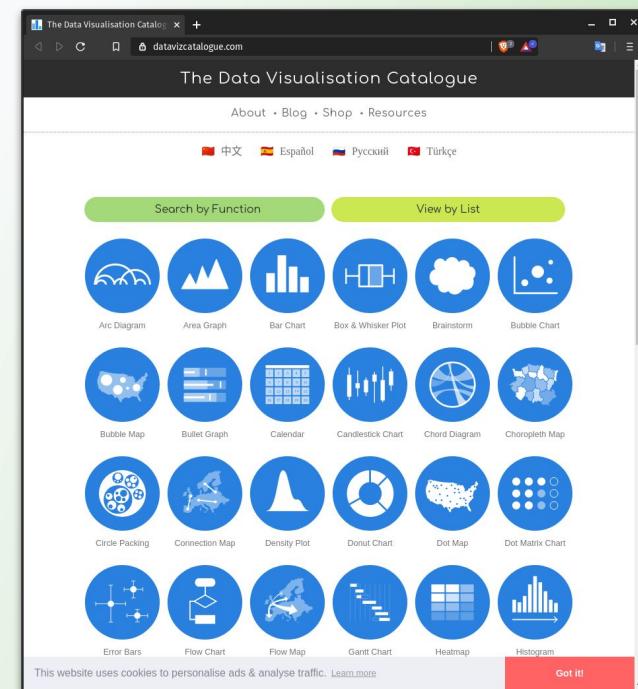
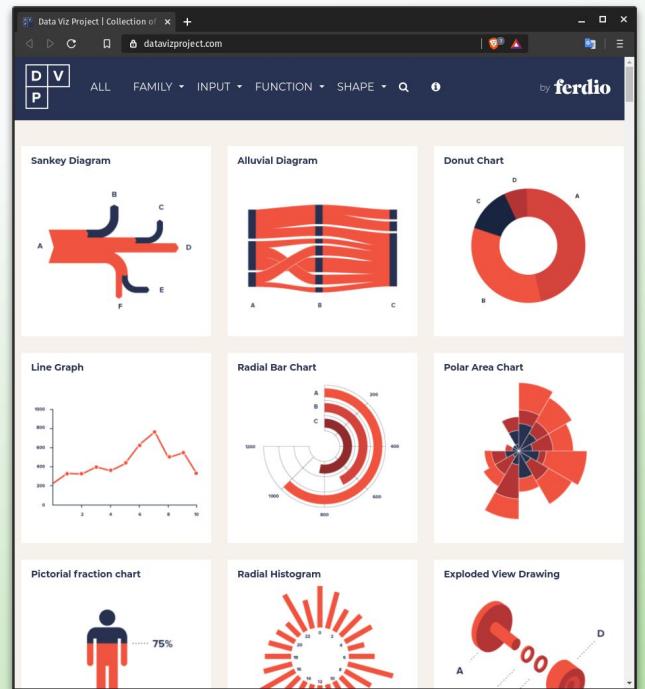
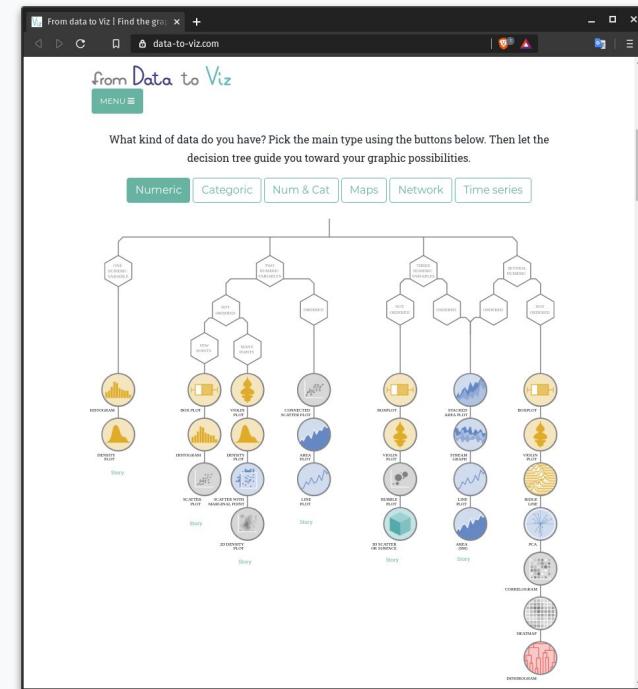
<https://bnhr.xyz>

What visualizations should I use

Data visualization catalogs

There are several free and open catalogs available online that will help you decide what chart to use based on your data.

Some even provide steps on how to create your charts.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Data visualization catalogs

- From Data to Viz: <https://www.data-to-viz.com/>
- Data Viz Project: <https://datavizproject.com/>
- The Data Visualisation Catalogue: <https://datavizcatalogue.com/>
- Chartmaker Directory: <http://chartmaker.visualisingdata.com/>
- Python Graph Gallery: <https://python-graph-gallery.com/>
- R Graph Gallery: <https://www.r-graph-gallery.com/>



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Activity - Data visualization catalogs

1. Visit any of these data visualization catalogs:
 - a. [The Data Visualisation Catalogue](#)
 - b. [From Data to Viz](#)
 - c. [Data Viz Project](#)
 - d. [Chartmaker Directory](#)
2. List down the visualizations that you want to try or caught your attention at this link: <https://bit.ly/3y8O2eY>



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

**When in doubt,
consult your audience.**



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Improving your visualizations



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

What can be improved in the table?

VACCINATION ROLLOUT IN ASEAN
(based on Bloomberg and reports from Foreign Service Posts,
as of 22 April 2021)

Rank	Countries	Vaccine Administered	Status of Vaccination	Doses Administered
1	Indonesia	Sinovac, AstraZeneca	Started 13-Jan-21	17,644,225
2	Singapore	Pfizer/BioNTech, Moderna	Started 30-Dec-20	2,213,888
3	Myanmar	AstraZeneca	Started 27-Jan-21	1,880,000
4	Philippines	Sinovac, AstraZeneca	Started 01-Mar-21	1,612,420*
5	Malaysia	Pfizer/BioNTech, Sinovac	Started 24-Feb-21	1,541,896
6	Cambodia	Sinopharm, AstraZeneca, Sinovac	Started 28-Feb-21	1,210,779
7	Thailand	AstraZeneca, Sinovac	Started 10-Feb-21	712,610
8	Laos	Sinopharm, Sputnik V	Started Feb-21	168,000
9	Vietnam	AstraZeneca	Started 08-Mar-21	106,929
10	Brunei	Sinopharm, AstraZeneca	Started 03-Apr-21	2,323

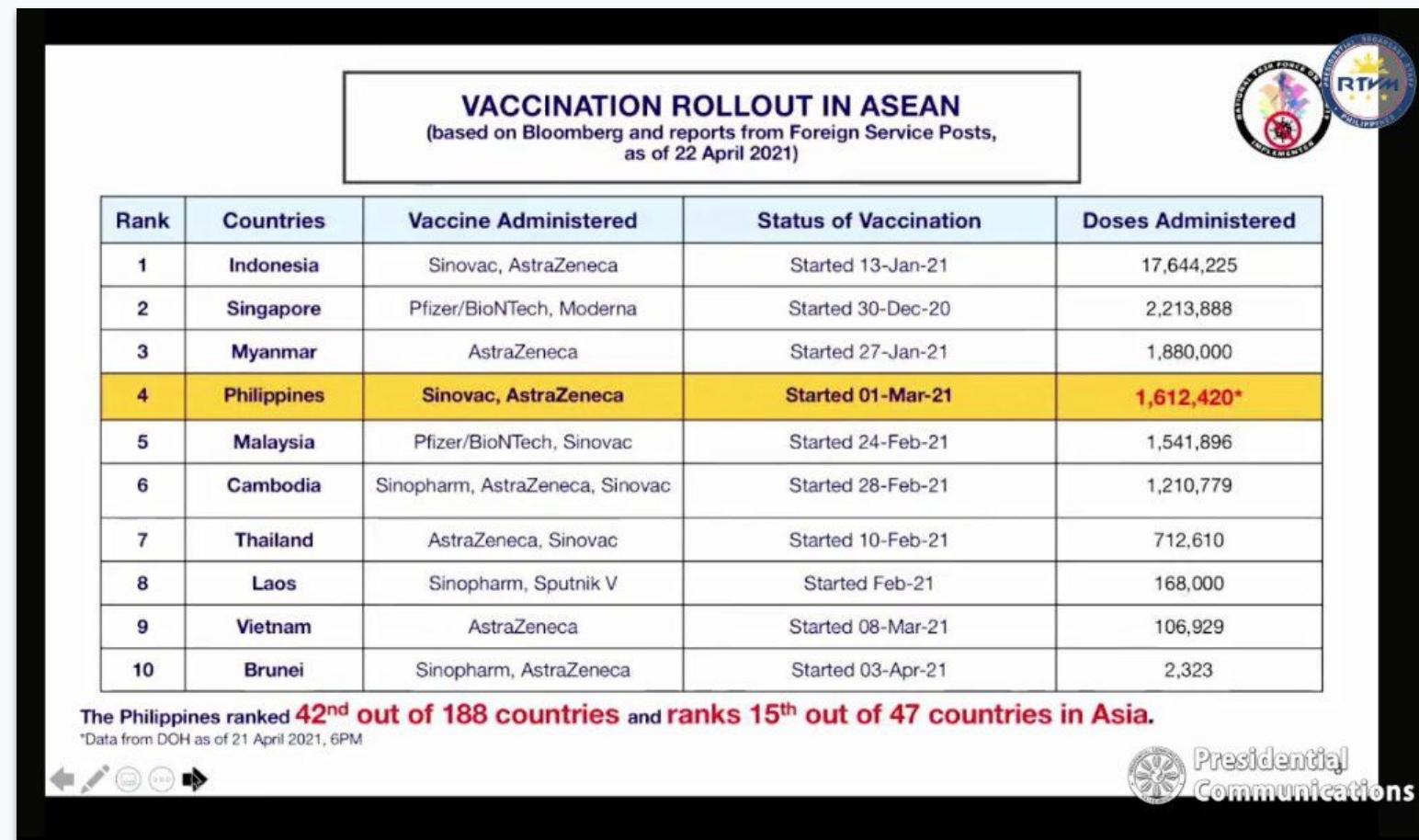
The Philippines ranked **42nd** out of 188 countries and ranks **15th** out of 47 countries in Asia.

*Data from DOH as of 21 April 2021, 6PM

← ↗ ↘ →

Presidential Communications

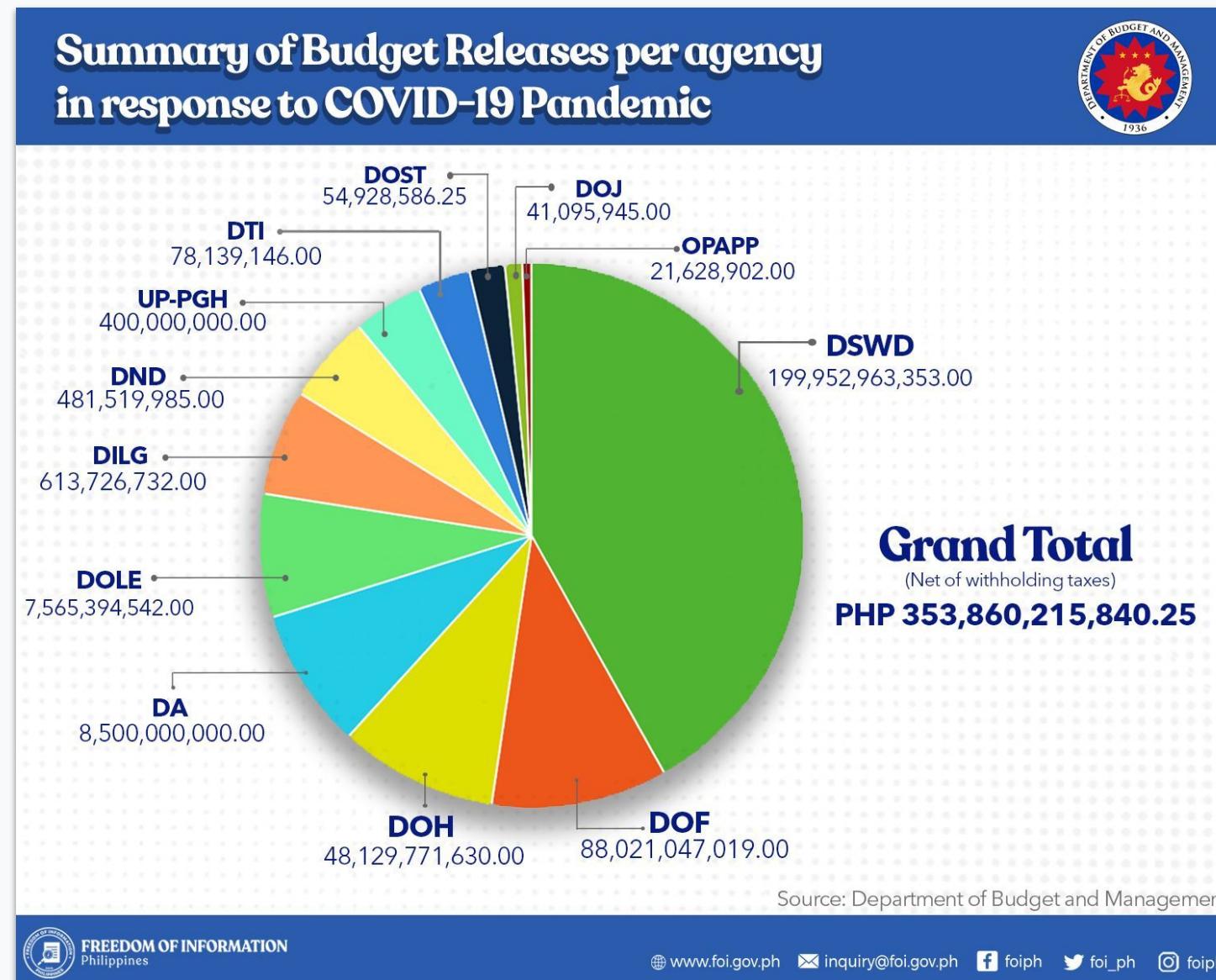
Normalize!



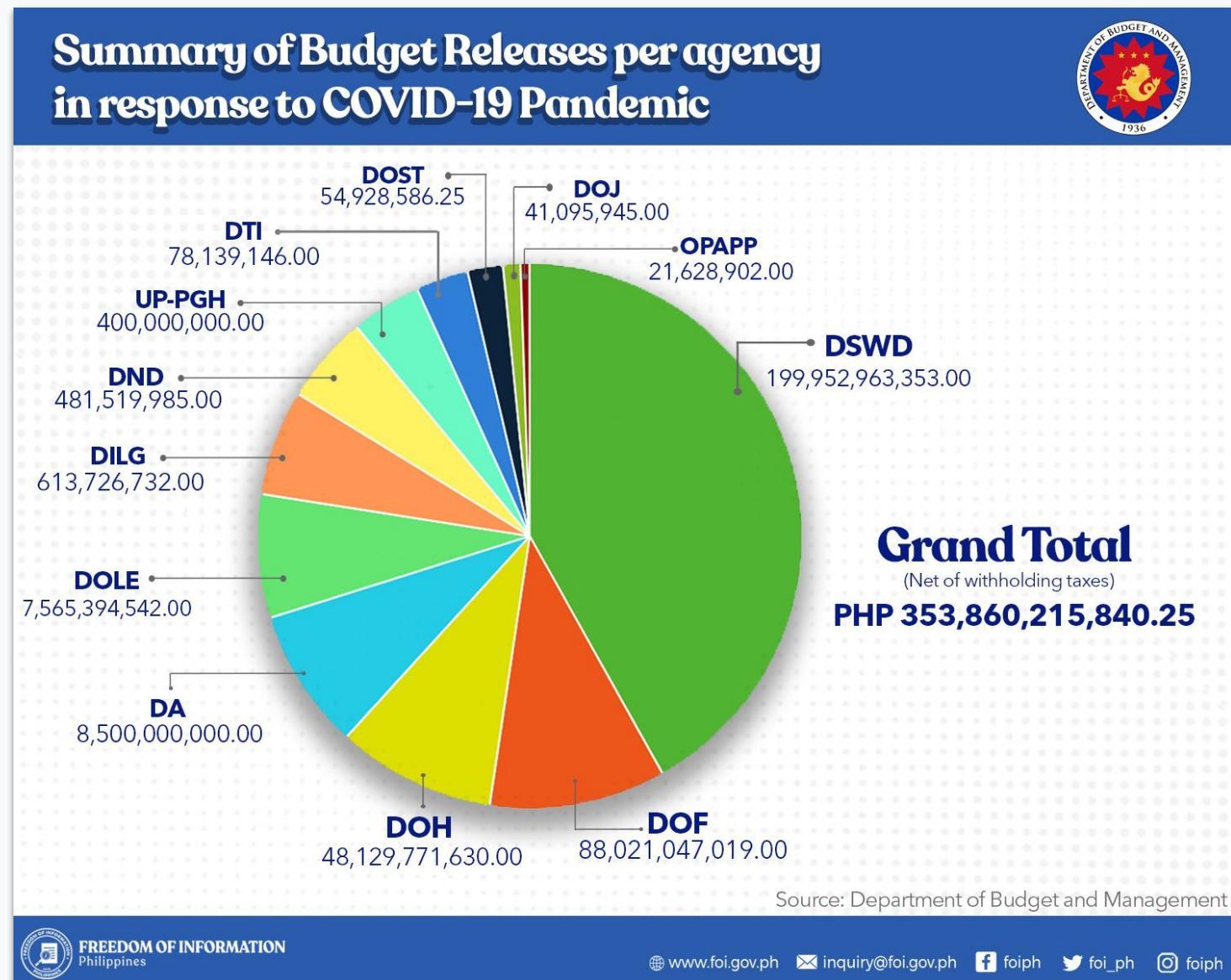
Doses are not normalized per population

- The ASEAN countries have different populations and the number of doses per population or percentage of population is a better metric/indicator.

What can be improved in the chart?



Look at the slices!



Too many slices

Slices are not proportional

- look at DOF and DOH
- look at DA and DOH

Improve your pie charts

Remove
to improve
the **pie chart** edition

Created by Darkhorse Analytics

www.darkhorseanalytics.com



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz

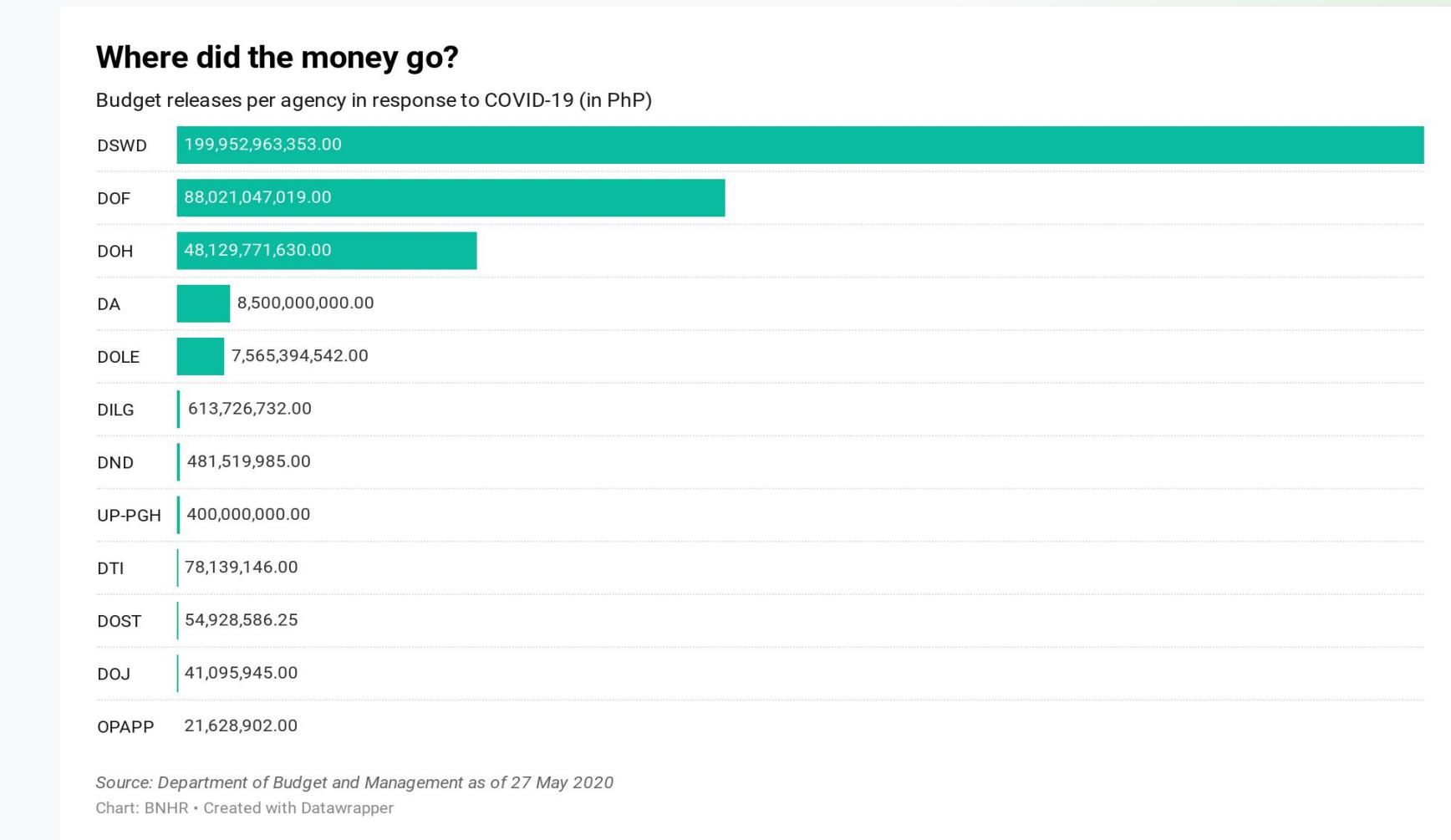
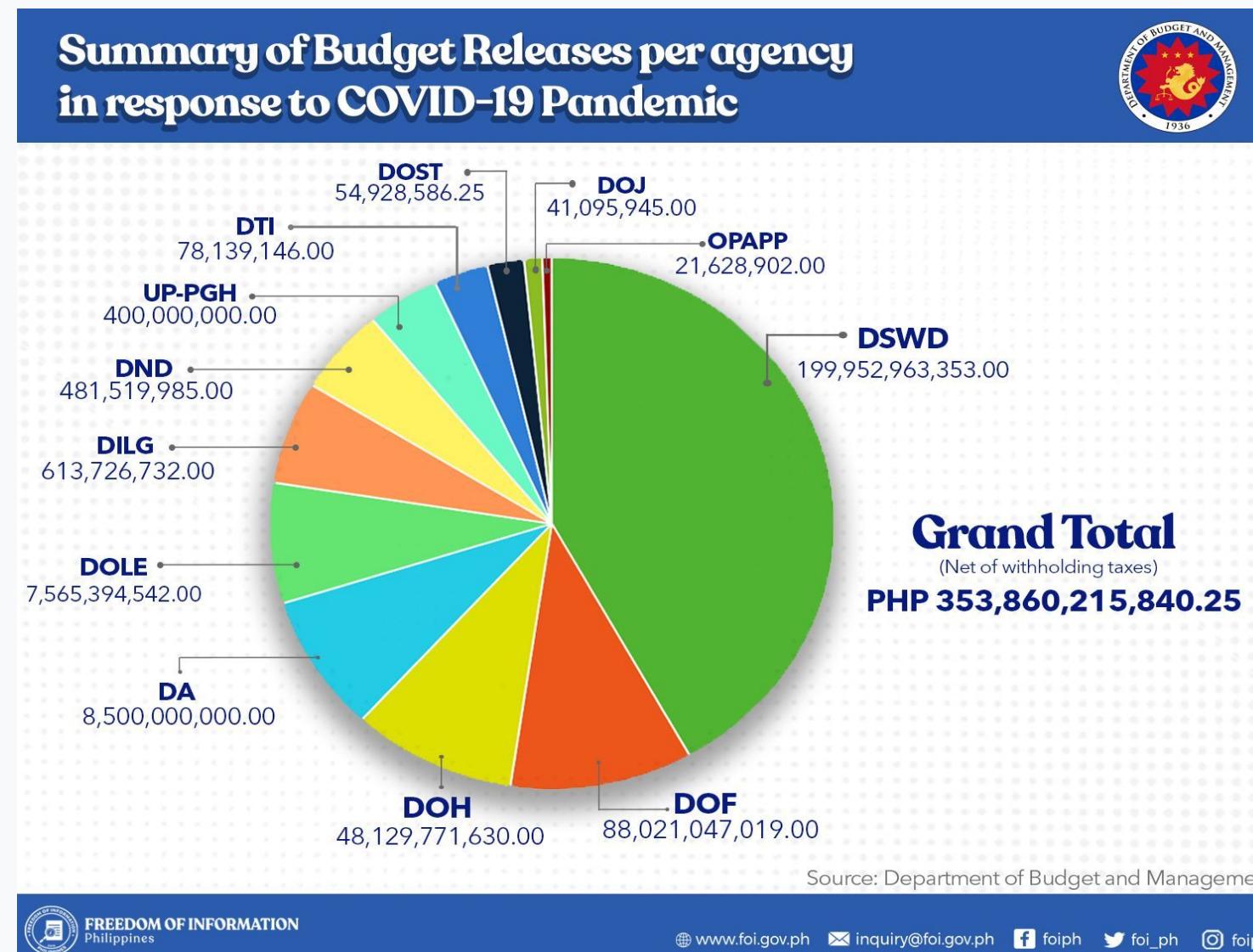


bnhrdotxyz



<https://bnhr.xyz>

Left: Original | Right: Modified



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz

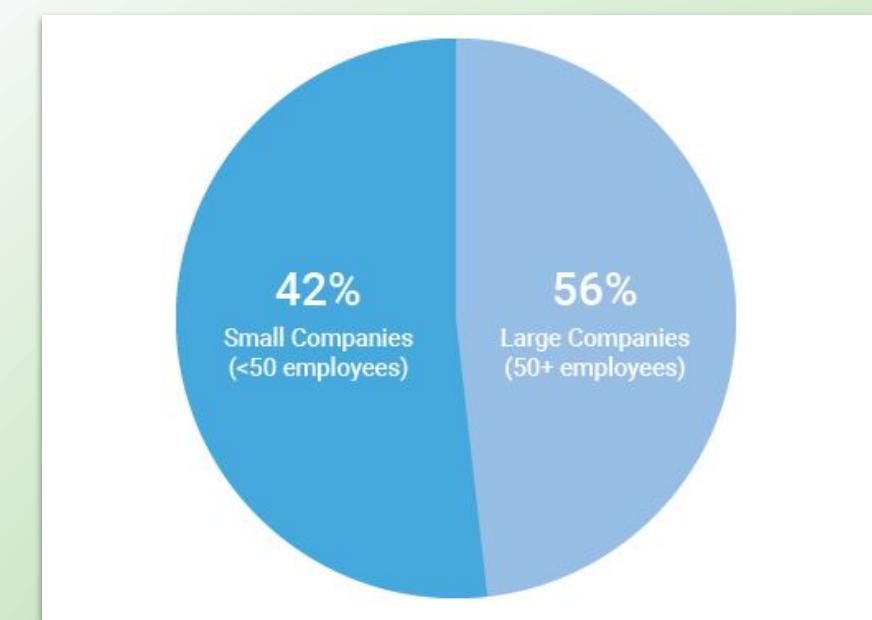
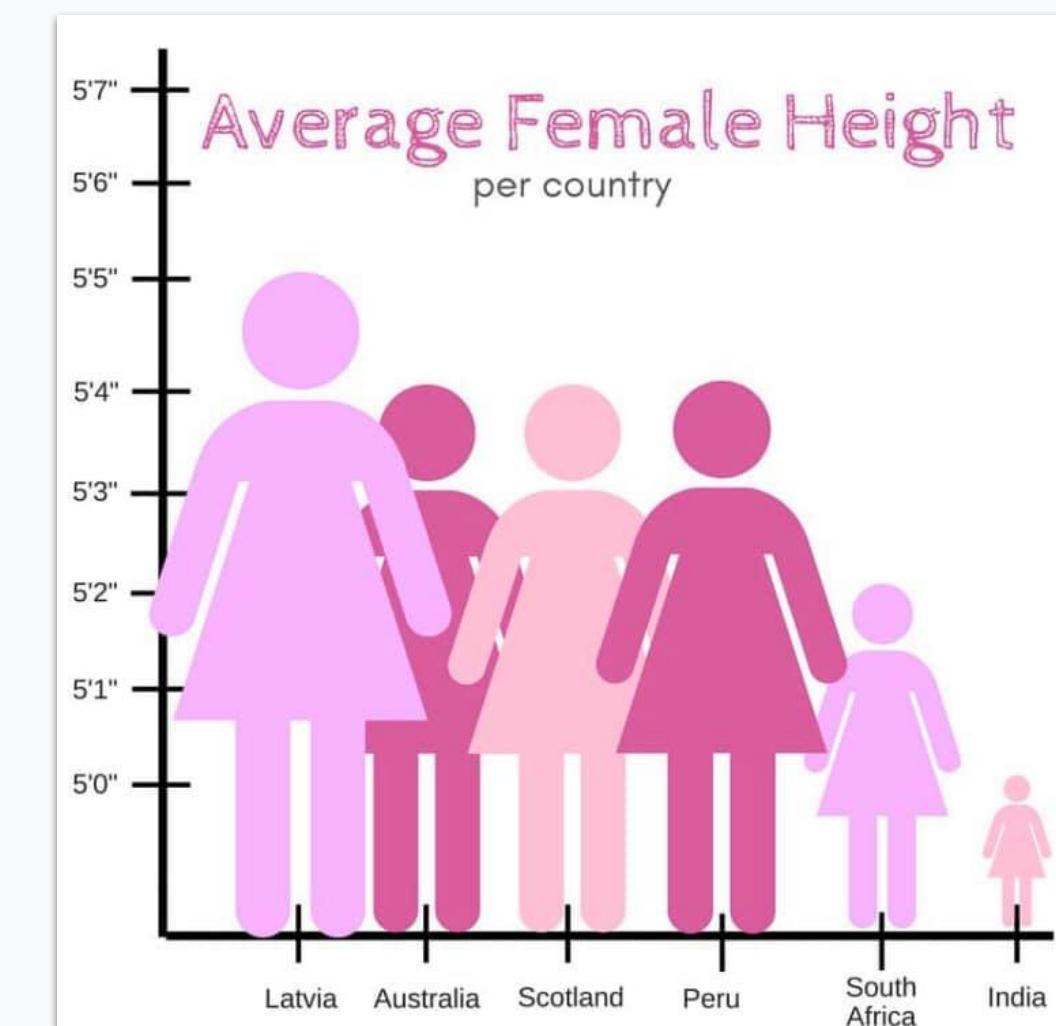
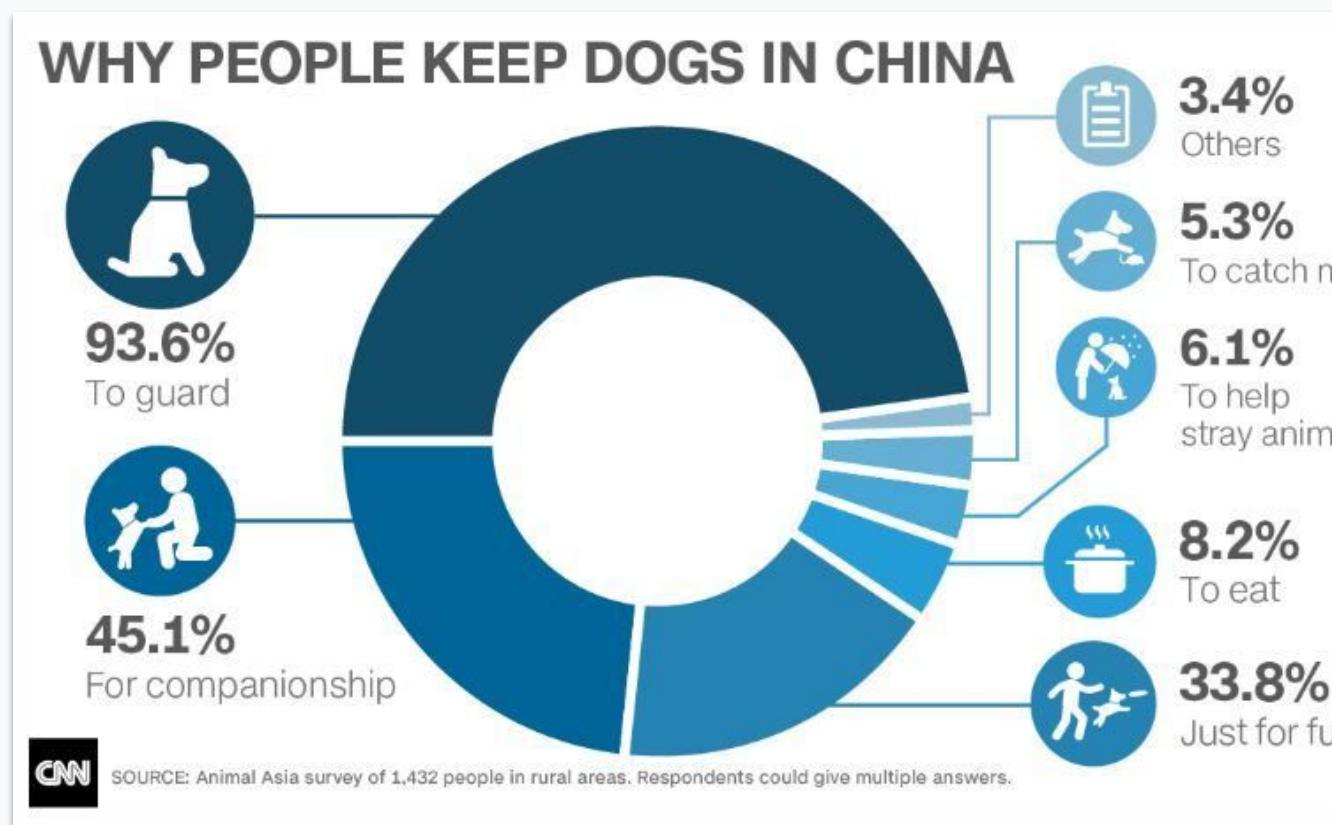


bnhrdotxyz

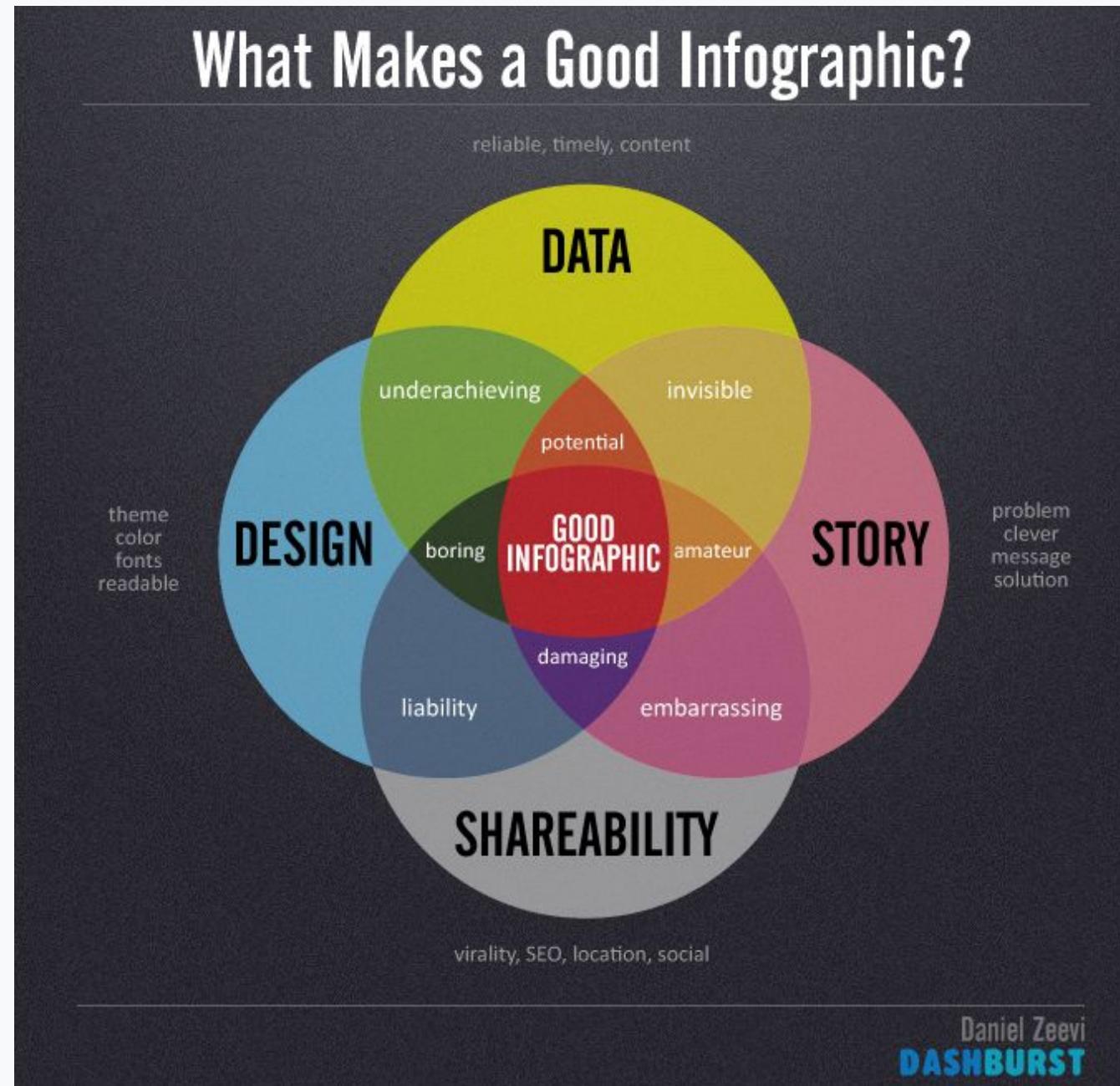


<https://bnhr.xyz>

What about these?



What makes a good infographic



What makes a good infographic by [Daniel Zeevi \(DASHBURST\)](#)

- **good data**
 - GIGO—garbage in garbage out
- **good design**
 - hooks audience and captures attention
 - choosing the right design elements
- **good story**
 - engages audience and keeps them reading
 - relatability and importance of story
- **shareable**
 - virality
 - what medium—physical? digital? social media?



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Design Principles



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Design principles to follow

- Simplification
- Hierarchy
- Layout



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Simplification

Focus only on
the things that matter



Created by Chameleon Design
from Noun Project



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz

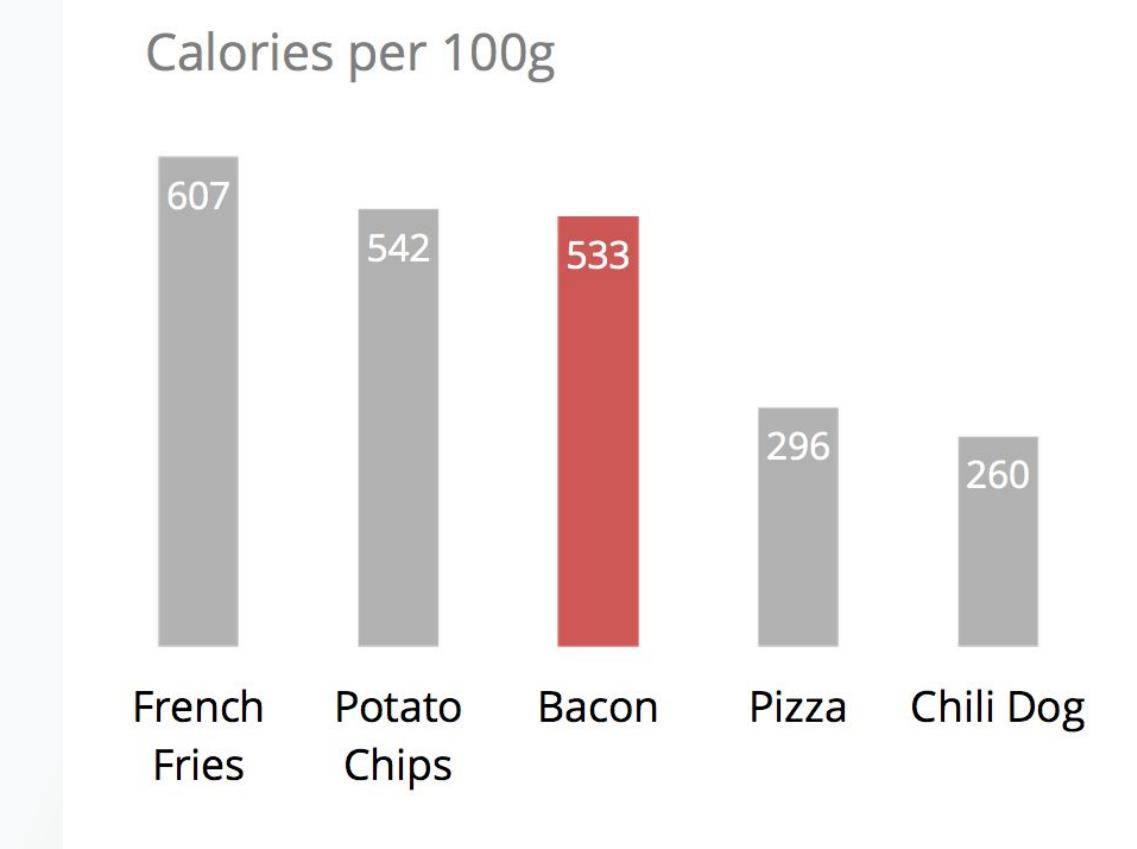
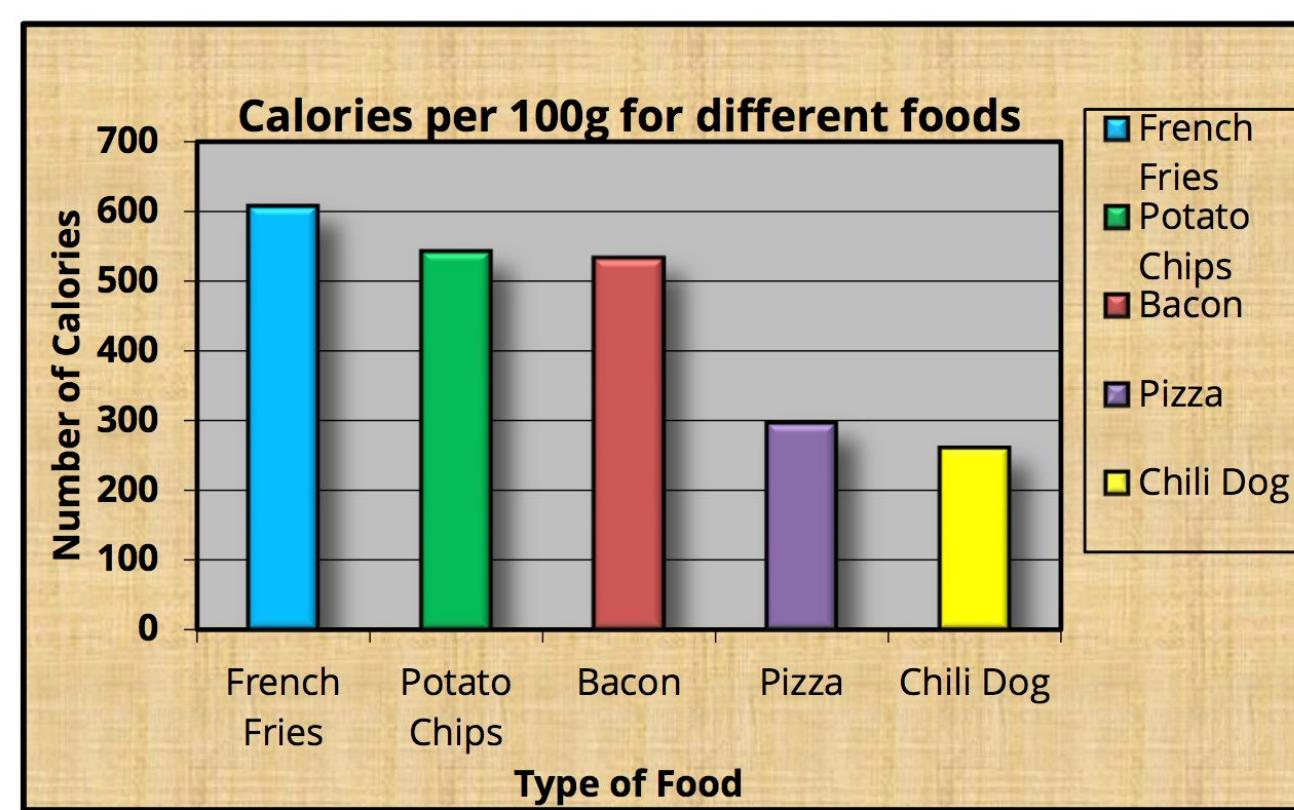


bnhrdotxyz



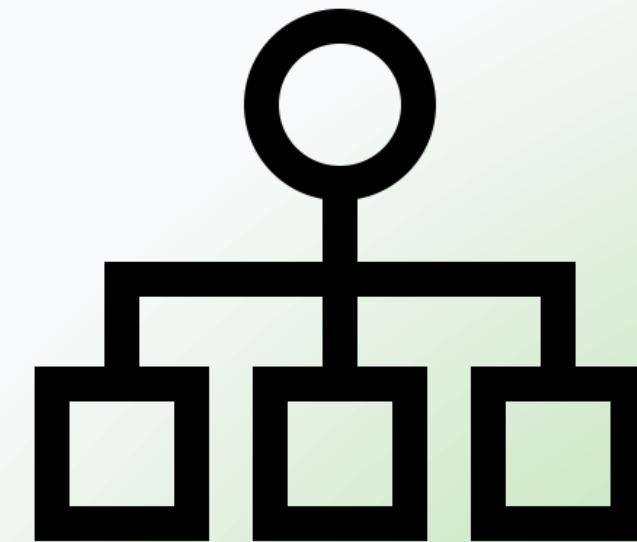
<https://bnhr.xyz>

Simplification



Hierarchy

Some objects are more important than others



Created by Dimitter Petrov
from Noun Project



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Hierarchy

- Utilize **size** - larger elements are more important than smaller elements.
- Utilize **contrast** - elements that pop off the page are more important than muted ones.
- Utilize **positioning** - our eyes usually follow a common pattern when reading a page, take advantage of that. All things being the same, people who read from left to right usually follow the “Z” path when looking at objects.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Size

This is probably more important

Than this but this is probably more important

Than this



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Contrast

This is probably something important

While this is probably less important

And this is probably just a side-note



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz

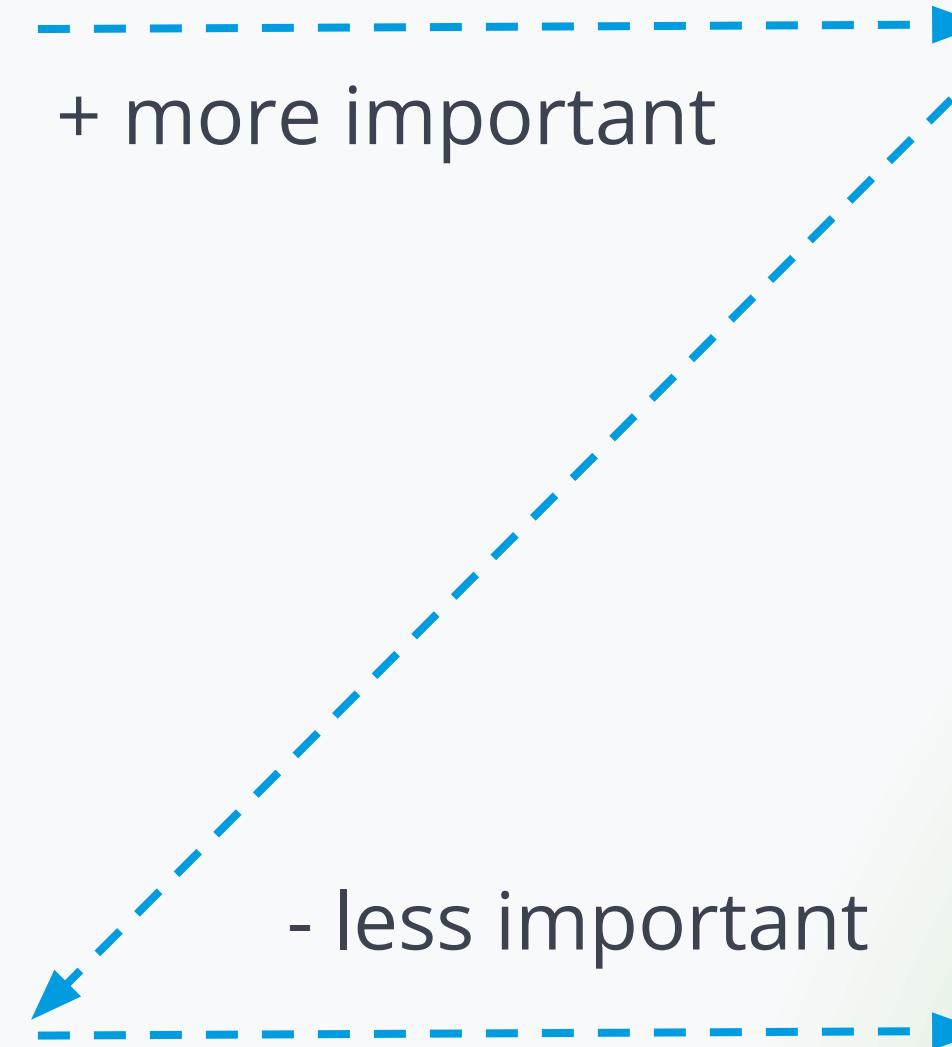


bnhrdotxyz



<https://bnhr.xyz>

Positioning



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Layout

Help users know where
and how to look



Created by joeartcon
from Noun Project



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Layout

- Portrait layout is good for documents and posters while a landscape is good for presentations.
- **Proximity** - elements placed close to each other will appear to be related.
- **Enclosure** - when you enclose things, you encourage the perception of these things belonging together.
- **Separation** - adding a line or space between elements makes them seem separated / disjoint.
- **Connection** - similar to enclosure, when you connect elements, you encourage the perception of these elements belonging to a group.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Design rules work

And you will read this last

You will read
this first

And then you will read this

Then this one



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Things to consider



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



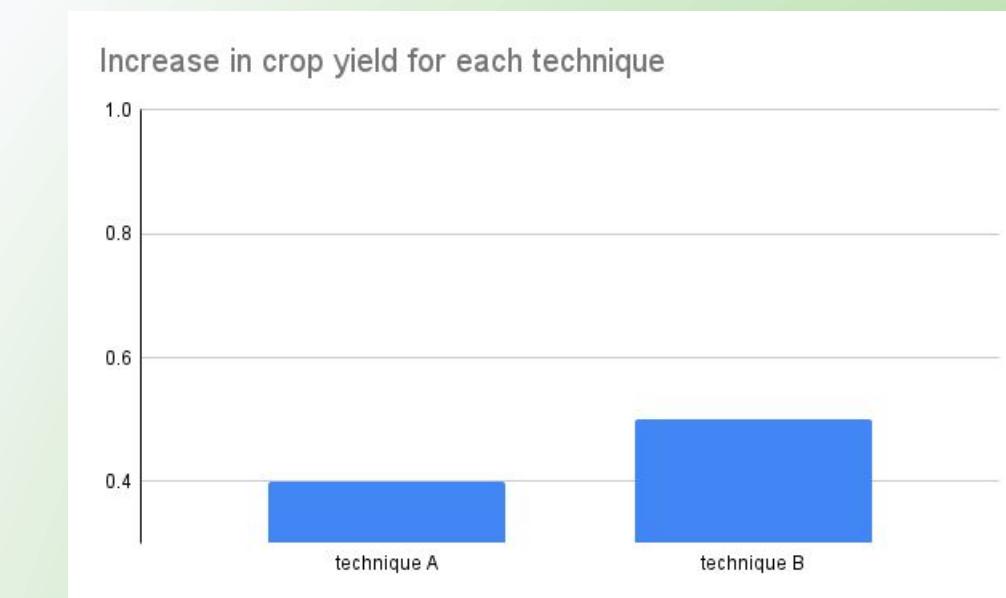
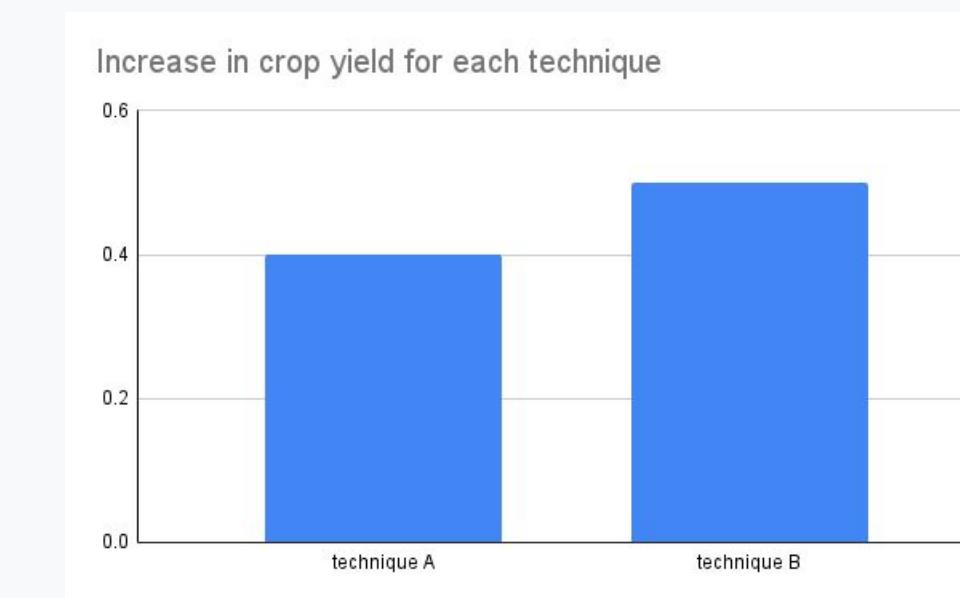
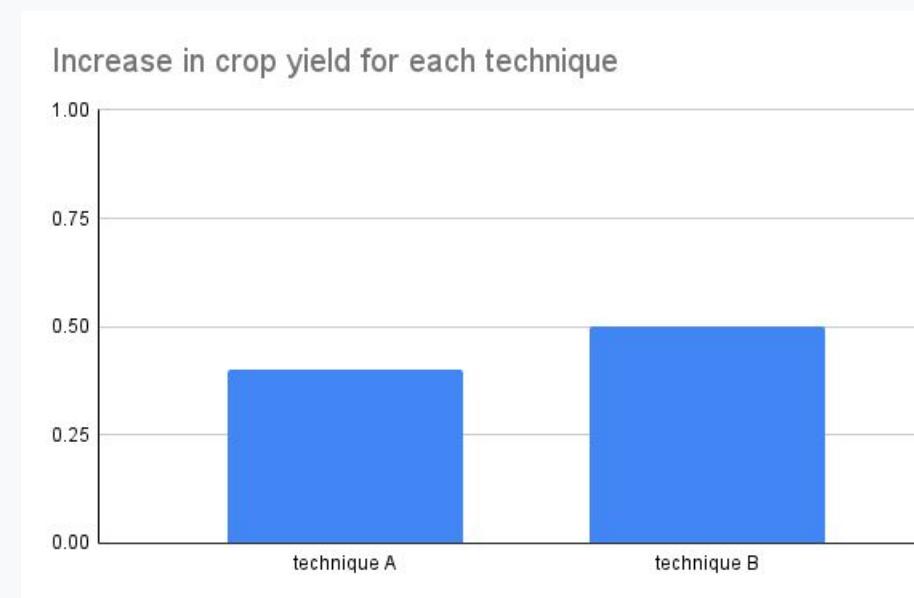
bnhrdotxyz



<https://bnhr.xyz>

Start at the beginning

- Begin bar charts at zero for a more accurate and truthful comparison of data.
- Truncating the Y-axis may lead to misrepresentation of the data that can mislead your audience.



Avoid distorting charts

- Line charts are great at showing trends but are sensitive to changes in the size of the chart.
- Stretching the height will tend to overplay (or overestimate) the changes or trend.
- Stretching the height will tend to underplay (or underestimate) the changes or trend.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Use pie charts sparingly

- Keep the number of categories to 5 or less.
- Don't use the pie for comparison—e.g. comparing the number of procurements per year
- Triple-check that the total value of your pie is exactly 100%.
- Don't make things worse by putting the pie in 3D.
- Pie charts are great at showing or highlighting one share of the total.
- Bar charts work great as substitutes for pie charts.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Use colors deliberately

- Use color for communication mostly and for decoration sparingly.
- Too many colors can confuse or disorient.
- Be mindful of your choices of color so you can create inclusive visualizations—e.g. color-blind friendly visulas, etc.
- There are online tools that can help you create color palettes and combinations that are inclusive, beautiful, and just work—e.g. Color Brewer, Viz Palette



liberty. data. geospatial.

 @bnhr.xyz

 @bnhrdotxyz

 bnhrdotxyz

 <https://bnhr.xyz>

Keep right

- Align whole numbers flush right to make it easy to find and compare them.
- Similarly, you can align decimals using their decimal point.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

More decimal places is not always better

- Extra decimal places imply more precise results but are usually unnecessary and distracting.
- Round your numbers off before plotting.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



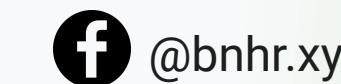
<https://bnhr.xyz>

Use both form and function

- Removing all artistic elements will make your visualization boring.
- Having too much artistic elements may distract the audience from the data and the message it conveys.
- Add labels and titles.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Don't use 3D if your data isn't 3D

- 3D is visually appealing but can make your chart hard to read and understand.
- Unless you are plotting something that is actually 3D (e.g. elevation), stick to 2D charts.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Not everything needs a chart

- Why use charts when you can use numbers directly?
- Why use charts when you can use bullet points?



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

There are always exceptions.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

**“If a rule prevents you from
improving something,
ignore it.”**

adapted from [Wikipedia](#)



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

**“By all means break the rules,
and break them beautifully,
deliberately, and well.”**

from [The Elements of Typographic Style](#)



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

“However, this concept of
‘breaking the rules’ is only
helpful if you already know
what you’re doing.”

from [What to consider when considering data vis rules](#)



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Free and open-source tools for working with data



liberty. data. geospatial.

 @bnhr.xyz

 @bnhrdotxyz

 bnhrdotxyz

 <https://bnhr.xyz>

Free, Open Source, Proprietary, Freeware

- **Free software**
 - free as in freedom ([Four Essential Freedoms of Free Software](#))
 - accessibility—i.e. no additional cost to download and install—is a fundamental characteristic
- **Open source**
 - under an [open source license](#)
 - often used synonymously with free software—being open source is a requirement for software to be free
- **Proprietary**
 - software that restrict what the user can do with the software (limited freedoms)—e.g. restricting access to source code, requiring payment of subscription, limits how the software may be used.
 - has a proprietary license
- **Freeware**
 - free as to cost
 - usually proprietary
 - freeware does not mean free software or open source



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Data collection



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



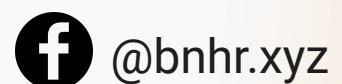
<https://bnhr.xyz>

ODK

What is it?	ODK is a suite of open source tools that help organizations collect and manage data consisting of ODK Collect, ODK Central, and ODK Build.
Free and open source	Yes
What can I do with it?	A common workflow using ODK: <ol style="list-style-type: none">1. Create survey forms using the XLSForm standard in Excel or Google Sheets.2. Upload forms to an ODK Central server.3. Download forms into ODK Collect on an Android device.4. Use Collect to fill out forms with participants.5. Upload survey data from Collect to Central.6. Analyze or export data from Central.
Ease of use	Easy to moderate depending on your needs and skills
Additional associated costs	Deployment of ODK Central if you will self-host Paid subscription is available Commercial support (training, consulting, etc.) is available
Where to learn more?	https://getodk.org/



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

ODK-X

What is it?	ODK-X is another free and open source toolkit to come out of the Open Data Kit project—the same one that created ODK. Compared to ODK, ODK-X is less mature but offers a more flexible and more complex tool suite that allows you to create your own data management applications.
Free and open source	Yes
What can I do with it?	<ul style="list-style-type: none">• ODK-X Survey- a customizable data collection application.• ODK-X Tables - a data curation and visualization application that can also run custom-built data collection workflows.• ODK-X Services - an application for user authentication and data synchronization between the ODK-X applications.• ODK-X Cloud Endpoints - a cloud server to host data and application files, and to support bi-directional data synchronization across mobile devices.• ODK-X Suitcase - a desktop tool for synchronizing data with a cloud endpoint.• ODK-X Application Designer - a design environment for creating, customizing, and previewing your forms, data curation, and visualization applications. This is where you build your ODK-X applications.
Ease of use	Easy to moderate depending on your needs and skills
Additional associated costs	Development and deployment of applications Commercial support (training, consulting, etc.) is available
Where to learn more?	https://odk-x.org/



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

KoBo Toolbox

What is it?	KoBoToolbox is a suite of free and open source tools for field data collection for use in challenging environments.
Free and open source	Yes
What can I do with it?	<ul style="list-style-type: none">• Design forms quickly and easily using a Form Builder• Collect data online and offline via an Android app (KoBoCollect) or on any modern web browser (Enketo).• Analyze and manage data through summary reports, graphs, tables, and maps.
Ease of use	Easy to moderate depending on your needs and skills
Additional associated costs	Deployment of server if you will self-host Commercial support (training, consulting, etc.) is available <i>* KoBo Toolbox maintains publicly-available instances:</i> <ul style="list-style-type: none">• <i>Humanitarian Server - Unlimited Use for Humanitarian Organizations</i>• <i>Non-Humanitarian Server - For Researchers, Aid Workers & Everyone Else</i>
Where to learn more?	https://www.kobotoolbox.org/



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

QField

What is it?	QField is a free and open source application that allows you to bring your QGIS project on the field. It works together with QFieldCloud and QFieldSync so that you can efficiently work on your GIS data outdoors.
Free and open source	Yes
What can I do with it?	<ol style="list-style-type: none">1. Create a project in QGIS on your desktop.2. Collect the data needed for your project on the field.3. Sync, edit, and manage the data you collected back into QGIS.
Ease of use	Easy to moderate depending on your needs and skills
Additional associated costs	Deployment of QFieldCloud if you will self-host Commercial support (training, consulting, etc.) is available
Where to learn more?	https://qfield.org/



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Mergin Maps

What is it?	Mergin Maps allows you to capture geospatial information easily through your mobile or tablet then share it with your team for seamless collaboration. It consists of the QGIS Mergin plugin, Mergin Maps Cloud, and Mergin Maps input.
Free and open source	Yes
What can I do with it?	<ol style="list-style-type: none">1. Setup your GIS project on QGIS2. Capture the data you need easily on mobile or tablet using the Mergin Maps Input app for iOS and Android3. Store and track changes to your geodata4. Integrate the data you collected into QGIS
Ease of use	Easy to moderate depending on your needs and skills
Additional associated costs	Deployment of Mergin Cloud if you will self-host Paid subscription is available Commercial support (training, consulting, etc.) is available
Where to learn more?	https://merginmaps.com/



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Data analysis and visualization



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

LibreOffice Calc

What is it?	LibreOffice Calc is the spreadsheet application of the free and open source office suite LibreOffice. A modern spreadsheet application with features that are as good as, if not better, than those of proprietary alternatives.
Free and open source	Yes
Do I need to install it?	Yes
Do I need an account?	No
What can I do with it?	<ul style="list-style-type: none">• Manage spreadsheets and other tabular data files• Perform computations and analyses• Create charts and graphs
Ease of use	Easy
Additional associated costs	Commercial support (training, consulting, etc.) is available
Where to learn more?	https://www.libreoffice.org/



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Airtable

What is it?	Airtable combines the ease-of-use of spreadsheets and the power of a database to provide a flexible framework for connecting data and building applications based on your data.
Free and open source	Free as to cost
Do I need to install it?	No
Do I need an account?	Yes
What can I do with it?	<ul style="list-style-type: none">• Manage tabular data (known as a Base)• Create applications such as forms and websites based on your Base• Perform computations and make visualizations usind your data• Share your data online
Ease of use	Depends on what you intend to use it for.
Additional associated costs	Free and paid options are available Commercial support (training, consulting, etc.) is available
Where to learn more?	https://www.airtable.com/



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Datawrapper

What is it?	Airtable combines the ease-of-use of spreadsheets and the power of a database to provide a flexible framework for connecting data and building applications based on your data.
Free and open source	Yes
Do I need to install it?	No but you can deploy your own instance
Do I need an account?	No
What can I do with it?	<ul style="list-style-type: none">• Create interactive and responsive charts (19 types)• Create interactive and responsive maps (choropleth map, symbol map, locator map)• Create interactive and responsive data tables• Share, download, or embed your charts, maps, and tables
Ease of use	Easy
Additional associated costs	Self-hosting costs Free and paid options are available Commercial support (training, consulting, etc.) is available
Where to learn more?	https://datawrapper.de



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

RAWGraphs

What is it?	RAWGraphs is an open source data visualization framework built with the goal of making the visual representation of complex data easy for everyone. Almost 30 visual models to visualize quantities, hierarchies, time series and find insights in your data.
Free and open source	Yes
Do I need to install it?	No but you can deploy your own instance
Do I need an account?	No
What can I do with it?	<ul style="list-style-type: none">• Create a wide range of charts from your data• Export your chart as a vector or raster image that you can edit in other softwa
Ease of use	Easy
Additional associated costs	Self-hosting costs Commercial support (training, consulting, etc.) is available
Where to learn more?	https://www.rawgraphs.io/



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

QGIS

What is it?	QGIS is a mature, cross-platform, free and open source Geospatial Information System (GIS) with features for collecting, storing, analysing, presenting, and managing spatial & non-spatial data and runs on GNU/Linux, macOS, Windows, and even Android. Integrates well with other existing geospatial technologies and serves as an integral part of any FOSS4G (Free and Open Source Software for Geospatial) stack.
Free and open source	Yes
Do I need to install it?	Yes
Do I need an account?	No
What can I do with it?	<ul style="list-style-type: none">• Manage, analyze, and visualize data in a spatial way• Create maps• Perform different kinds of spatial analysis such as: find the service areas of public utilities, find the shortest path between points, find suitable vaccination sites, determine flood risk
Ease of use	Depends on what you intend to use it for.
Additional associated costs	Commercial support (training, consulting, etc.) is available
Where to learn more?	https://qgis.org/



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Leaflet

What is it?	Leaflet is the leading open-source JavaScript library for mobile-friendly interactive maps.
Free and open source	Yes
Do I need an account?	No
What can I do with it?	<ul style="list-style-type: none">• Create web/mobile maps and geospatial web/mobile applications.
Ease of use	Depends on what you intend to use it for.
Additional associated costs	Commercial support (training, consulting, etc.) is available
Where to learn more?	https://leafletjs.com/



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

MapLibre

What is it?	MapLibre is a community-governed collection of open source mapping libraries. The initial libraries are forks of the Mapbox GL ecosystem—for the web and mobile platforms—that maintain their open source license.
Free and open source	Yes
Do I need an account?	No
What can I do with it?	<ul style="list-style-type: none">• Create web/mobile maps and geospatial web/mobile applications.
Ease of use	Depends on what you intend to use it for.
Additional associated costs	Commercial support (training, consulting, etc.) is available
Where to learn more?	https://maplibre.org



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Mapbox

What is it?	Mapbox is a collection of products that allow you use location information to create applications and solutions.
Free and open source	Free as to cost, Source-available
Do I need an account?	Yes
What can I do with it?	<ul style="list-style-type: none">• Create web/mobile maps and geospatial web/mobile applications.
Ease of use	Depends on what you intend to use it for.
Additional associated costs	Paid subscription is available Commercial support (training, consulting, etc.) is available
Where to learn more?	https://mapbox.com



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

OpenStreetMap

What is it?	OpenStreetMap is a map of the world, created by people and free to use under an open license.
Free and open source	Yes
Do I need an account?	Yes for adding edits. No for using the map/data.
What can I do with it?	<ul style="list-style-type: none">• Add location data (roads, buildings, points of interest) on OSM as open data• Use the data stored in OSM (in spatial analysis, in your app, as basemaps, etc.)• Used a lot in humanitarian settings like disaster preparedness (Humanitarian OpenStreetMap Team, Missing Maps)• Used in a lot of advocacy work (MapBeks, Mental Health Awareness Mapping, Community Pantry Mapping, etc.)
Ease of use	Depends on what you intend to use it for.
Additional associated costs	Commercial support (training, consulting, etc.) is available
Where to learn more?	https://openstreetmap.org



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Other tools (proprietary)

- Google Sheets
- Canva
- Tableau
- Flourish
- Google Data Studio
- PowerBI



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Activity - Trying out some tools

Try out <https://app.datawrapper.de> and <https://www.rawgraphs.io/> on some data.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>

Thank you!



openness.
data.
geospatial.

 @bnhr.xyz

 @bnhrdotxyz

 @bnhrdotxyz

 bnhr.xyz



Except when expressly provided, this work and its contents by [Ben Hur S. Pintor](#) is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#).

Other works (images, media, software, etc.) referenced in this work are under their own respective licenses.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.



liberty. data. geospatial.



@bnhr.xyz



@bnhrdotxyz



bnhrdotxyz



<https://bnhr.xyz>