



CPSC 100

Computational Thinking

Data Visualization

Instructor: Firas Moosvi
Department of Computer Science
University of British Columbia



Agenda

- Project Reminders
- Topic switch (Data Visualization this week, Data provenance next week)
- Data Visualization!

Course Admin



Course Admin

- **Project Milestone 2 feedback released last week!**
 - Some of you needed to resubmit, TAs are looking at the resubmissions now - **do not wait for approval**, start working on Milestone 3!
 - Visit TA student hours for feedback
- **Project Milestone 3 is due Nov. 19th at 6 PM**
 - Note that in Labs this week there will be an activity where you can share your ideas and drafts with others and get feedback.
 - Please make sure you have ***something*** to share with others!



CPSC 100

Computational Thinking

Intro to Visualization

Instructor: Firas Moosvi
Department of Computer Science
University of British Columbia

Learning Goals



Learning Goals

After this **today's lecture**, you should be able to:

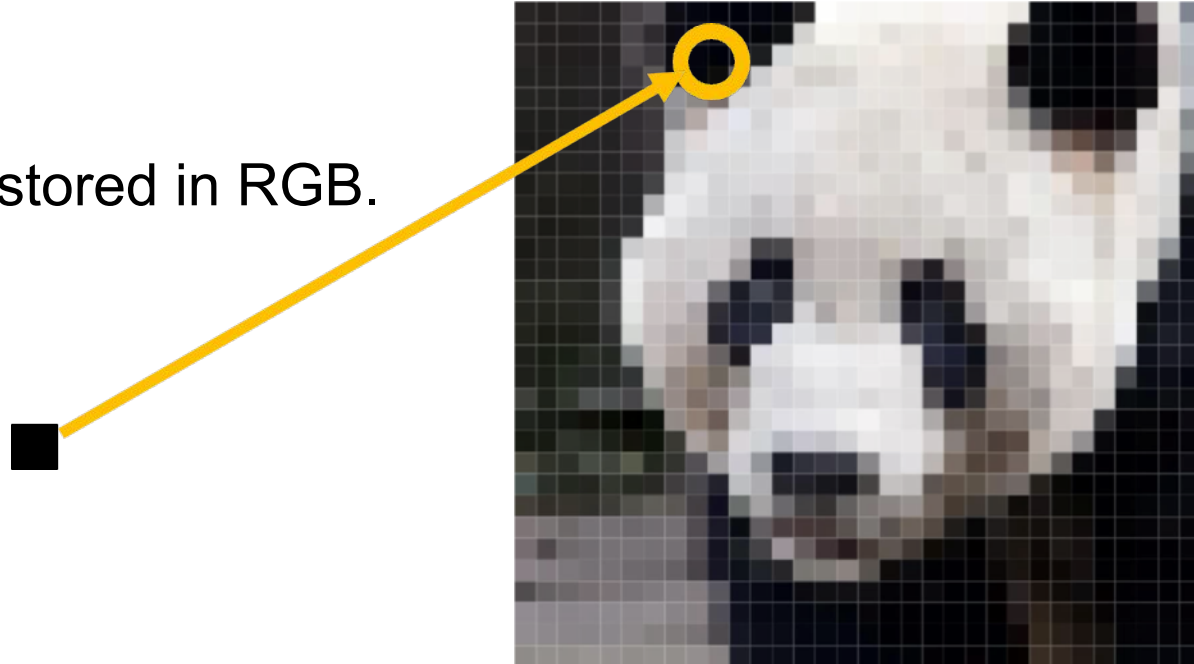
- Describe how images are stored in computers
- Explain the concept of bitmap images & how they are divided into pixels.
- Define visualization and explain its role in computational thinking.
 - Describe how visual data representation helps in understanding patterns, trends, and insights.
 - Differentiate between explanatory and exploratory visualization.
- Create different types of visual representations for a provided dataset.

Pixels & Computer



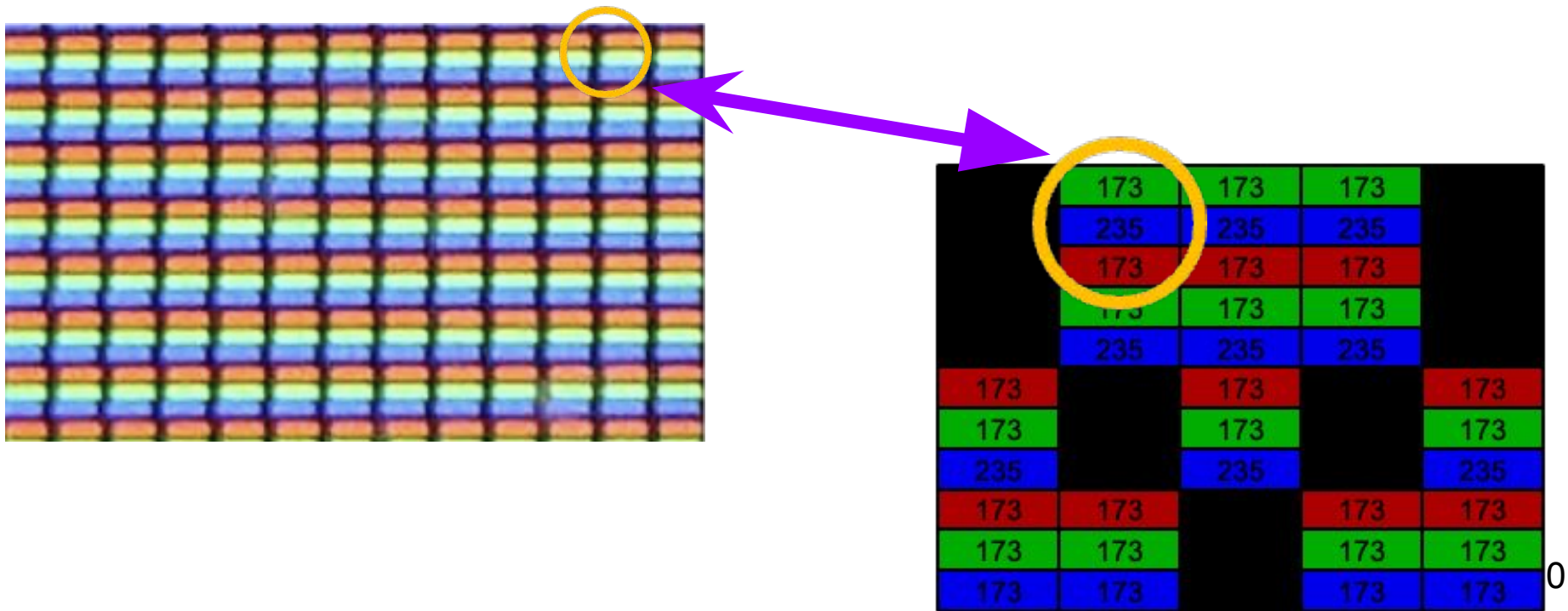
How are Images Stored?

- There are two ways that computers commonly store images. The most common is a **bitmap** – the picture is chopped up into small little squares called **pixels**.
- Each pixel's colour is stored in RGB.



Bitmap Pixels

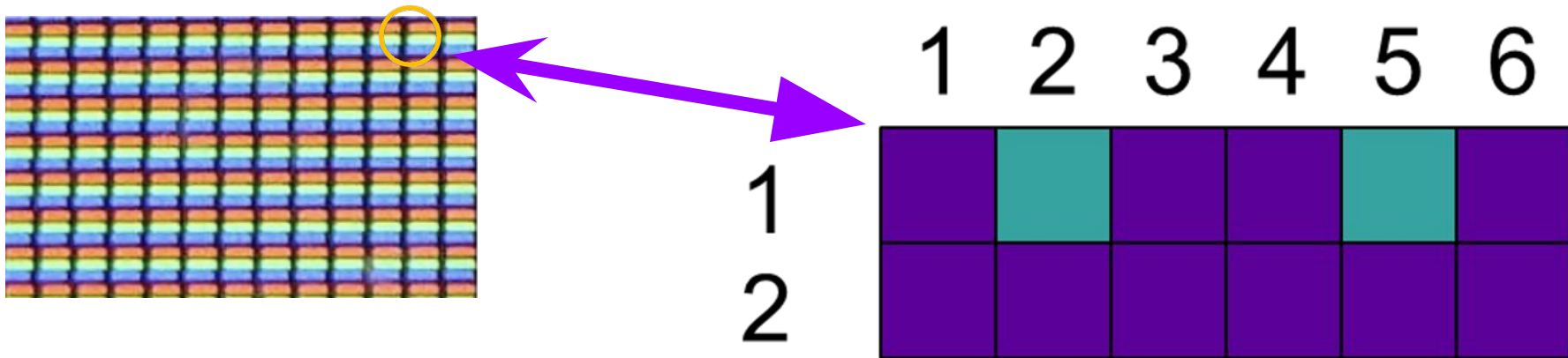
- The image is split into smaller **pixels**, each of which contain Red, Green and Blue





Bimap Pixels

- Bitmap image representation example



Note: the bitmap is

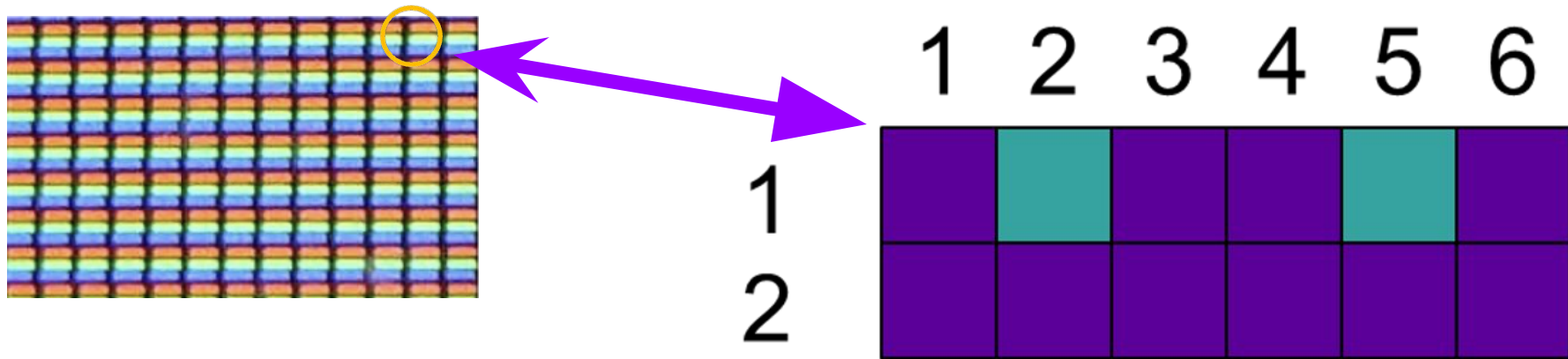
06 02

5C009A	35A4A0	5C009A	5C009A	35A4A0	5C009A
5C009A	5C009A	5C009A	5C009A	5C009A	5C009A



Bimap Pixels

- Bitmap image representation example



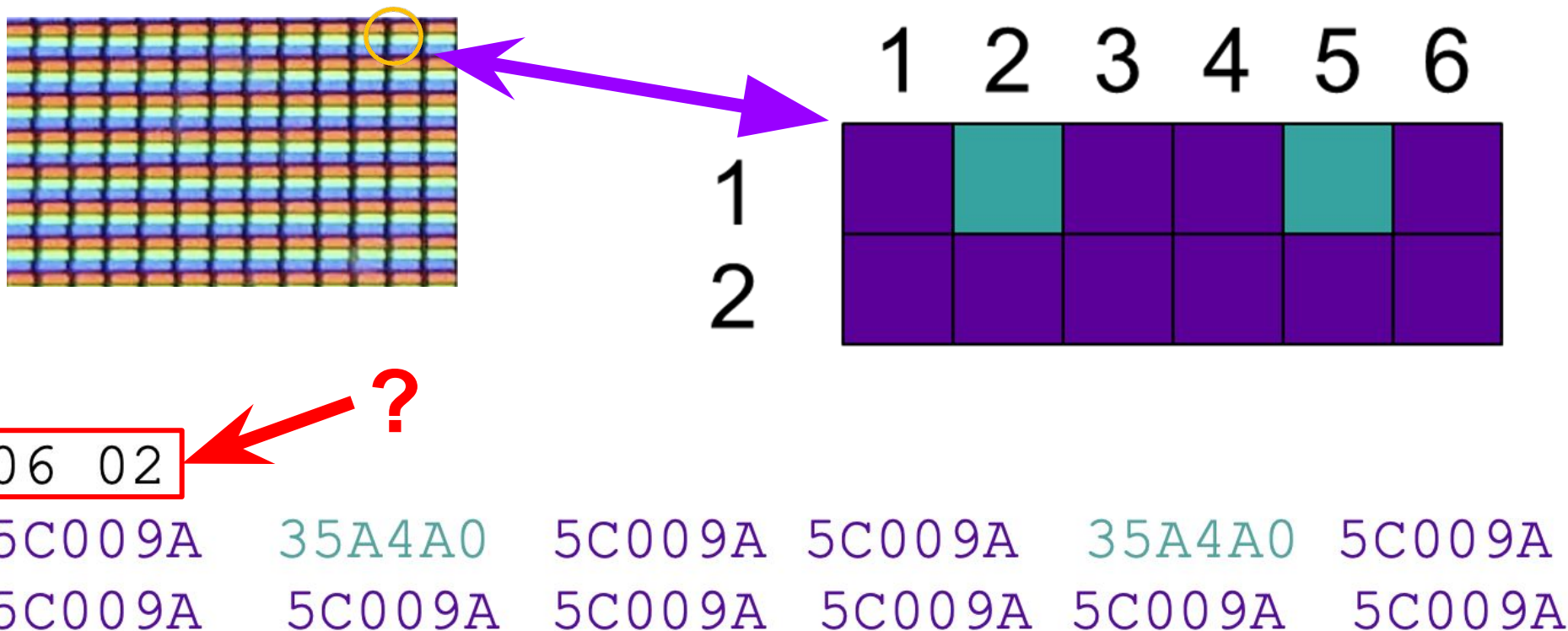
06 02

5C009A	35A4A0	5C009A	5C009A	35A4A0	5C009A
5C009A	5C009A	5C009A	5C009A	5C009A	5C009A



Bimap Pixels

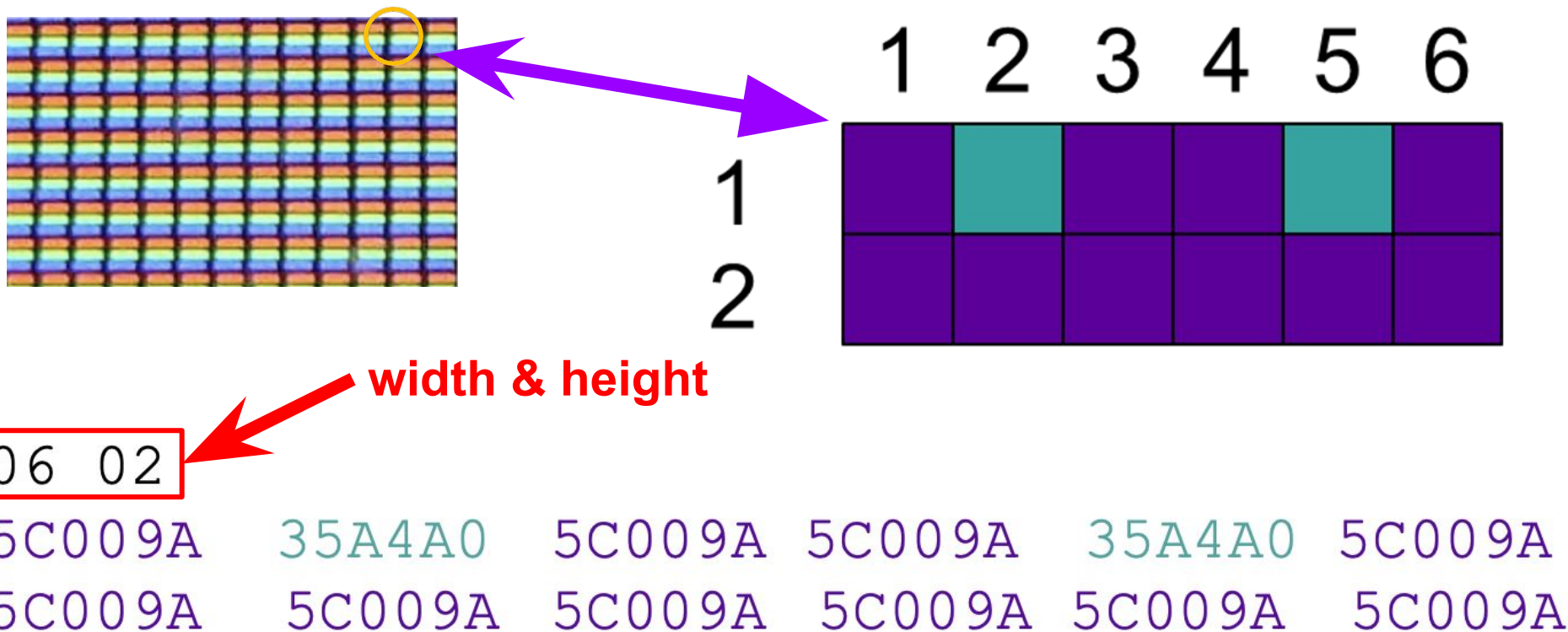
- Bitmap image representation example





Bimap Pixels

- Bitmap image representation example







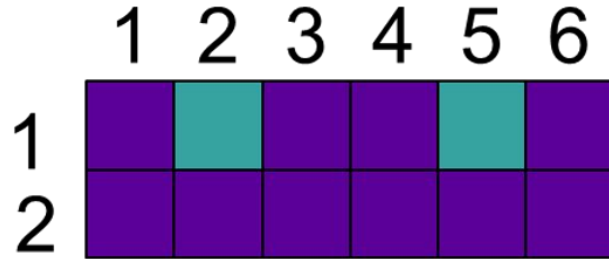
Q: Which image has the following bitmap?



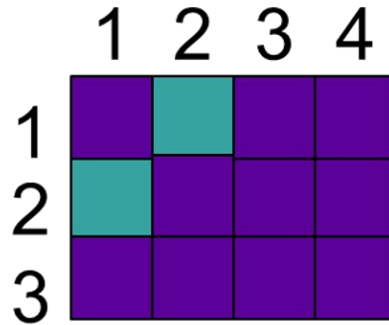
iClicker

04 03

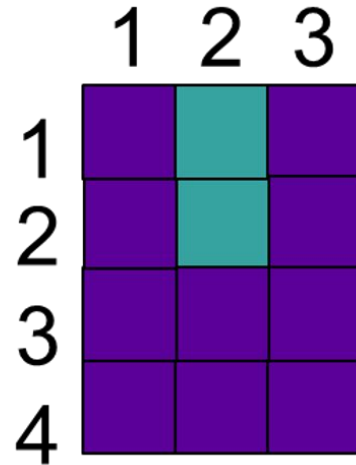
5C009A 35A4A0 5C009A 5C009A 35A4A0 5C009A
5C009A 5C009A 5C009A 5C009A 5C009A 5C009A



[A]



[B]



[C]



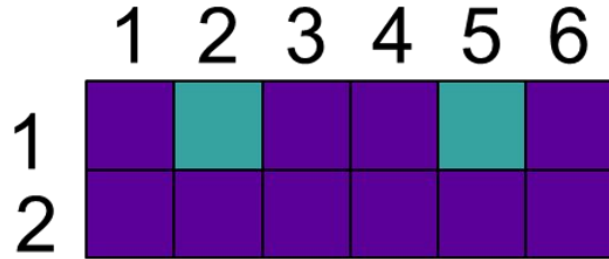
Q: Which image has the following bitmap?



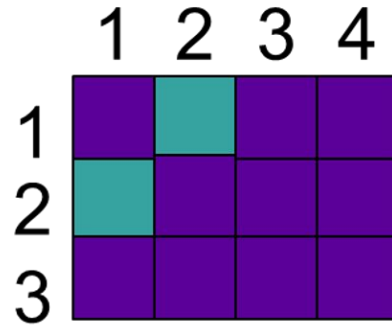
iClicker

04 03

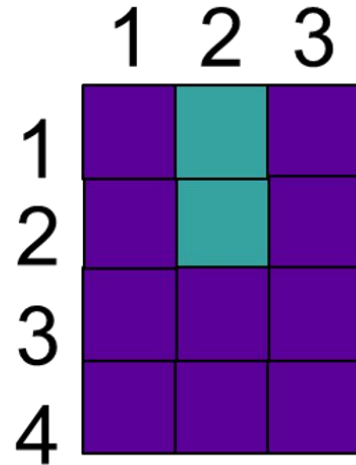
5C009A 35A4A0 5C009A 5C009A 35A4A0 5C009A
5C009A 5C009A 5C009A 5C009A 5C009A 5C009A



[A]



[B]



[C]

Data Visualization



What is Visualization?

- Also known as *visual data representation*
- The process of representing data graphically to make it easier to understand patterns, trends, and insights. Instead of raw numbers or complex tables, we use charts, graphs, and maps to communicate information effectively



What is Visualization?

- Also known as *visual data representation*
- The process of representing data graphically to make it easier to understand patterns, trends, and insights. Instead of raw numbers or complex tables, we use charts, graphs, and maps to communicate information effectively
- **Explanatory:** There is something in your data you would like to communicate to your audience
- **Exploratory:** You are trying to explore and understand patterns and trends within your data.

Why Visual Data Representation?



Why Visual Data Representation?

- Vision is our most dominant sense
- We are very good at recognizing visual patterns
- We need to see and understand in order to explain, reason, and make decisions:
 - Makes complex data easier to understand
 - Supports better decision-making
 - Improve communication



Why Visual Data Representation?

- Can you see any differences in the general trends of these four sets of numbers?

I		II		III		IV	
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Anscombe's Quartet

Four datasets that share the same descriptive statistics, including mean, variance, and correlation.

I

10	8.04
8	6.95
13	7.58
9	8.81
11	8.33
14	9.96
6	7.24
4	4.26
12	10.84
7	4.82
5	5.68

II

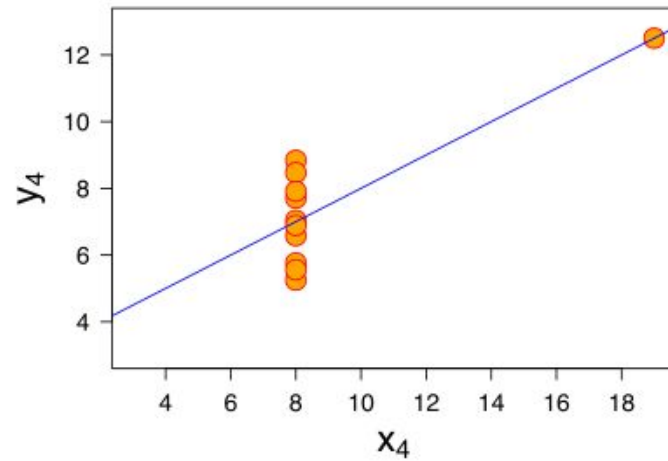
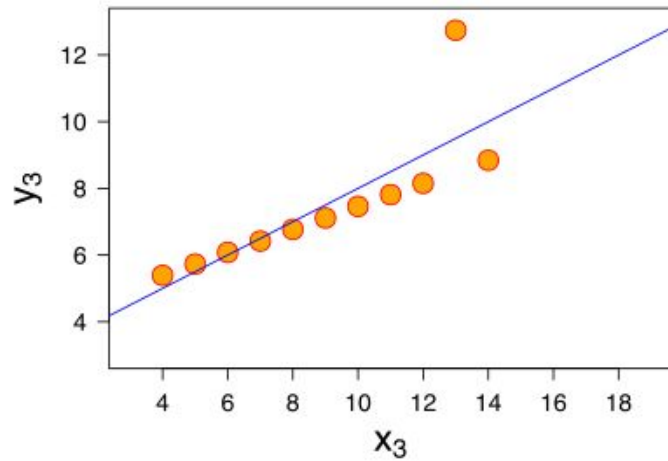
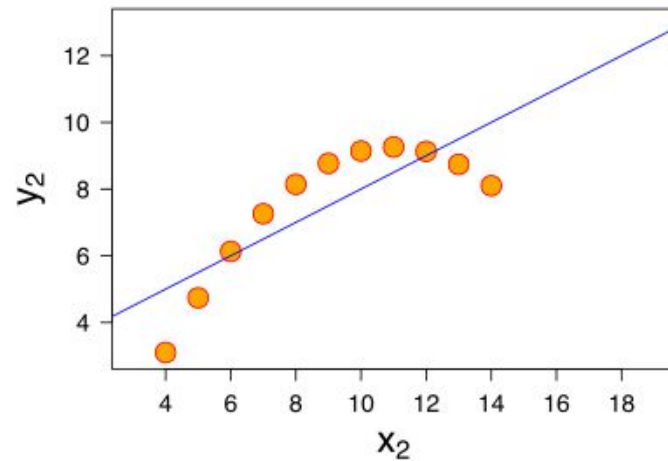
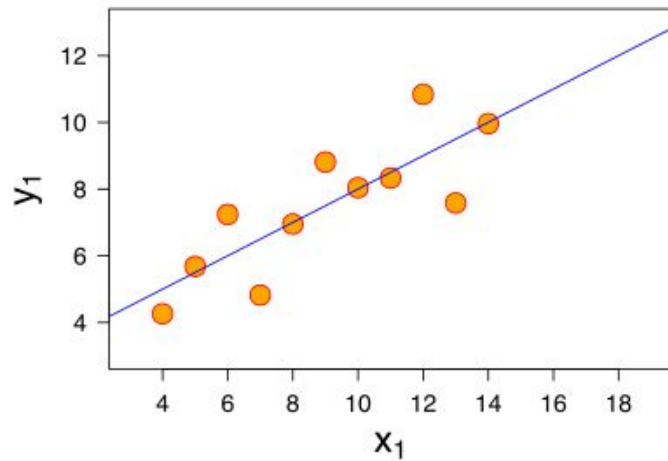
10	9.14
8	8.14
13	8.74
9	8.77
11	9.26
14	8.1
6	6.13
4	3.1
12	9.13
7	7.26
5	4.74

III

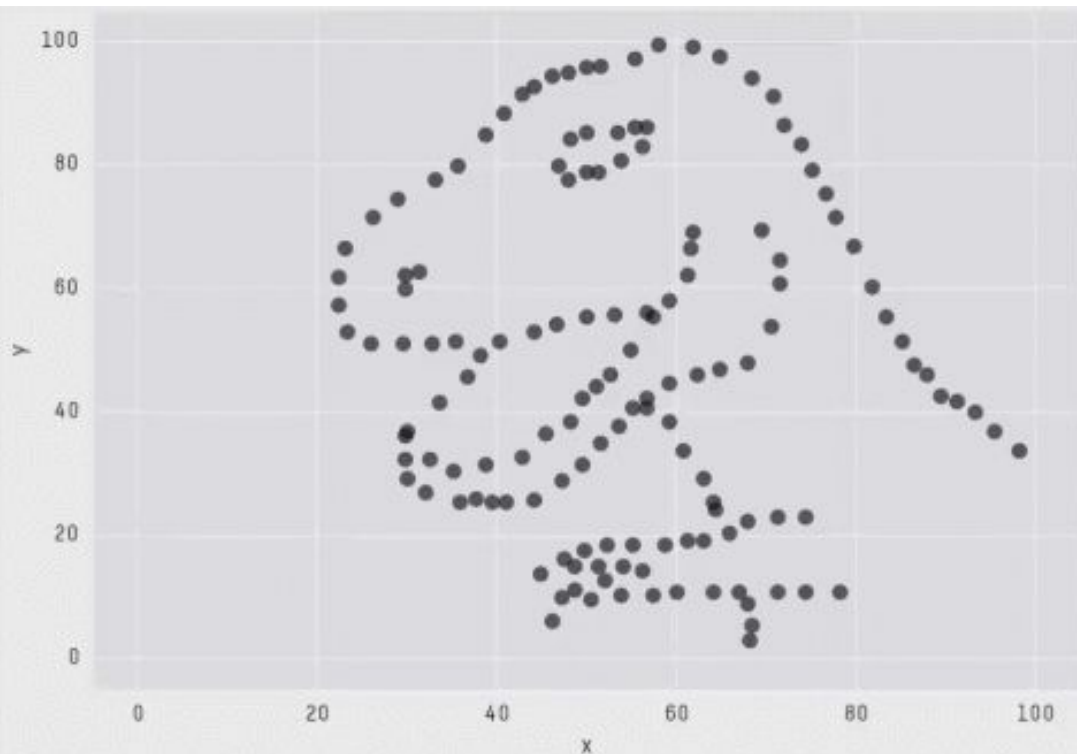
10	7.46
8	6.77
13	12.74
9	7.11
11	7.81
14	8.84
6	6.08
4	5.39
12	8.15
7	6.42
5	5.73

IV

8	6.58
8	5.76
8	7.71
8	8.84
8	8.47
8	7.04
8	5.25
19	12.5
8	5.56
8	7.91
8	6.89



Other Example: Datasaurus dataset [\[Link\]](#)



X Mean: 54.2659224
 Y Mean: 47.8313999
 X SD : 16.7649829
 Y SD : 26.9342120
 Corr. : -0.0642526

Activity



Activity

- Visualize the dataset below in at least 5 different ways.
- Sketch on paper
OR use a tool (e.g.Excel)

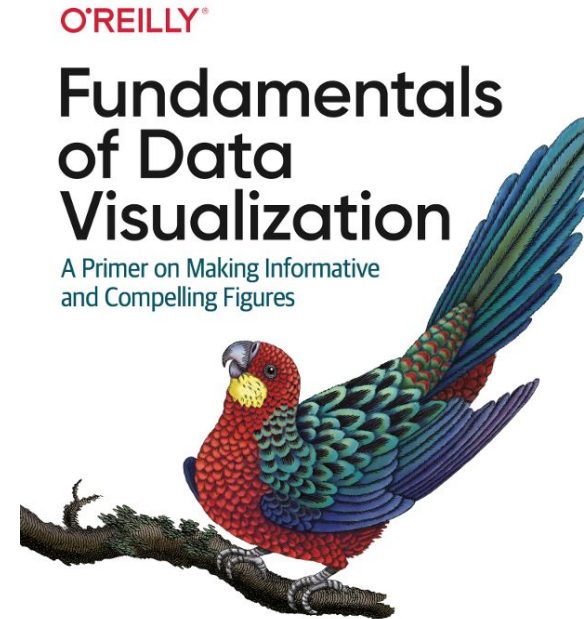
Ice Cream Flavour	Count
Chocolate	50
Vanilla	43
Mint	31
Strawberry	35
Matcha	12
Blueberry	43
Coffee	66

Debrief



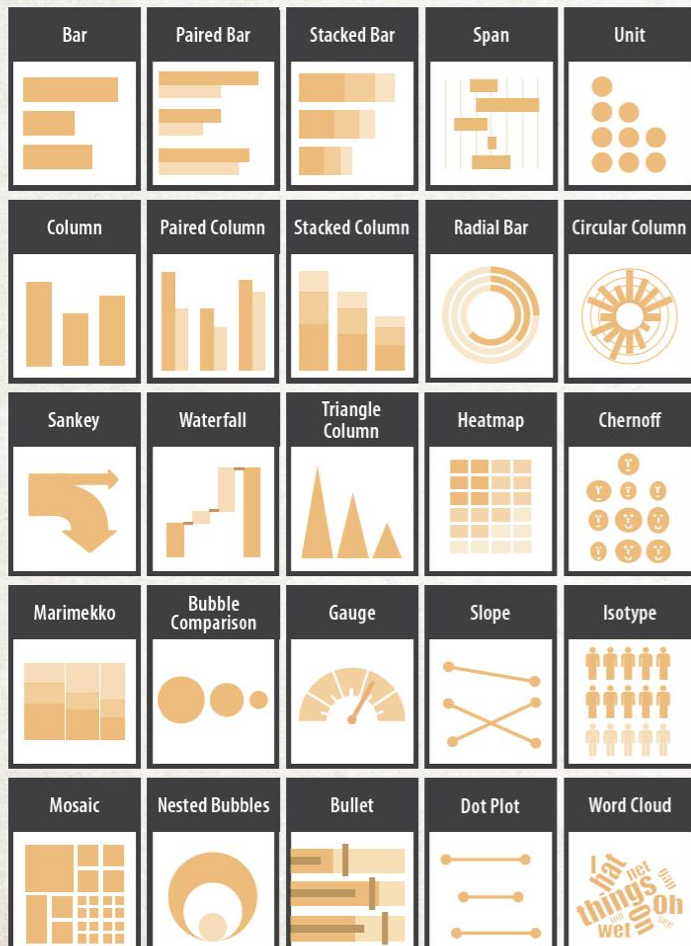
Welcome to the wonderful world of Data Visualization!

- Highly recommend this free online book as a reference!



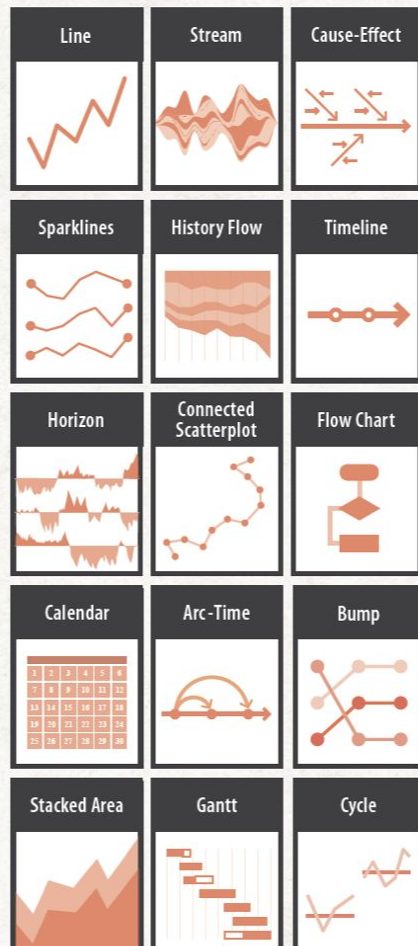
COMPARING CATEGORIES

Compare values across categories



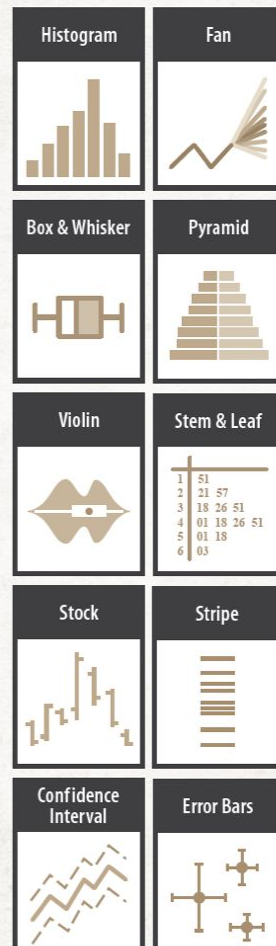
TIME

Track changes over time



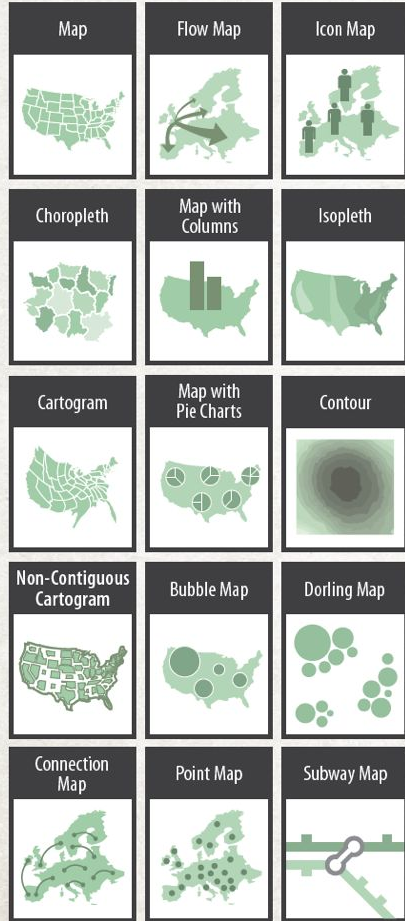
DISTRIBUTION

Representation of the distribution of data



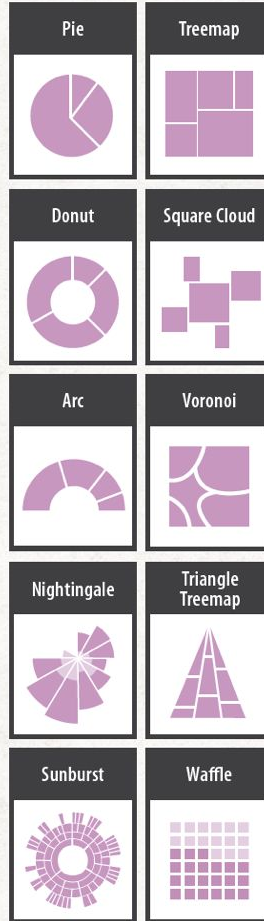
GEOSPATIAL

Relates data to its geography



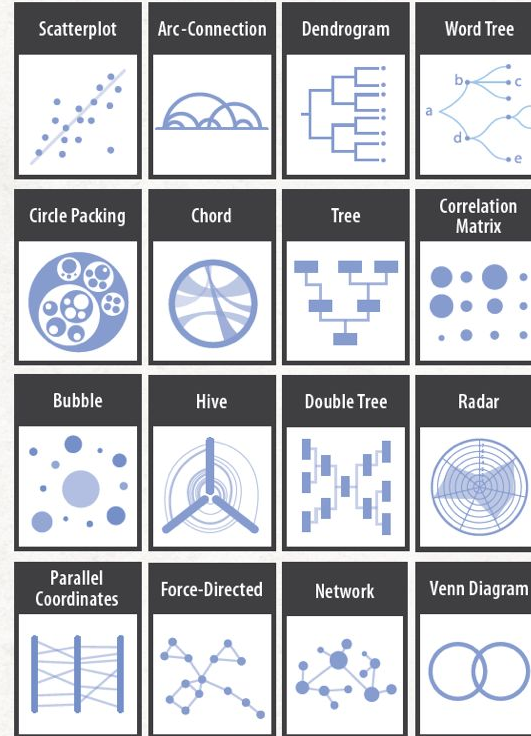
PART-TO-WHOLE

Relates the part of a variable to its total



RELATIONSHIP

Illustrates correlations or relationships between variables

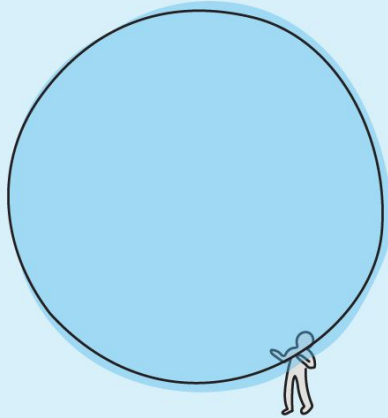


Principles of Effective Visualizations

Principle	Definition	Examples
• Proportional Ink	The amount of ink used to indicate a value should be proportional to the value itself.	Truncating the y-axis on a bar chart to exaggerate the difference between bars violates the principle of proportional ink.
• Data:ink ratio	Remove distracting visual elements to focus attention on the data	Lighten line weights, remove backgrounds, never use 3D or special effects, remove avoid unnecessary/redundant labels.
• Labels & legends	Use axes labels and titles to highlight/communicate data	Never leave your data column names as axes labels! Generally good to add a title.
• Overplotting	With large datasets, points overlap, resulting in large clouds of data	To fix overplotting, could plot just a sample subset of the data, use alpha, and use smaller points. Or, jitter - but check if appropriate!
• Visualization choice	Must be informed by the data you have, the research question being asked and the audience that cares.	Pick the simplest plot that best shows most/all of the data needed to answer the research question. If you only have summary statistics, cannot show distributions. Tailor the visualization to your audience (within reason) but don't dumb it down.
• Colour & Accessibility	Colour can be used to encode information or for aesthetics/style/design. However, colour can also be distracting if used inappropriately or poorly.	Choose a perceptually uniform colour palette; can be sequential or diverging for quantitative data. Opt for colour-blind friendly palettes. Categorical data can use qualitative colour schemes.



A REALLY HARD THING



HOW IT FEELS
RIGHT NOW



HOW IT WILL FEEL
IN A FEW MONTHS



HOW IT WILL FEEL
IN A FEW YEARS

LIZ FOSSLIE

Preview for next class...



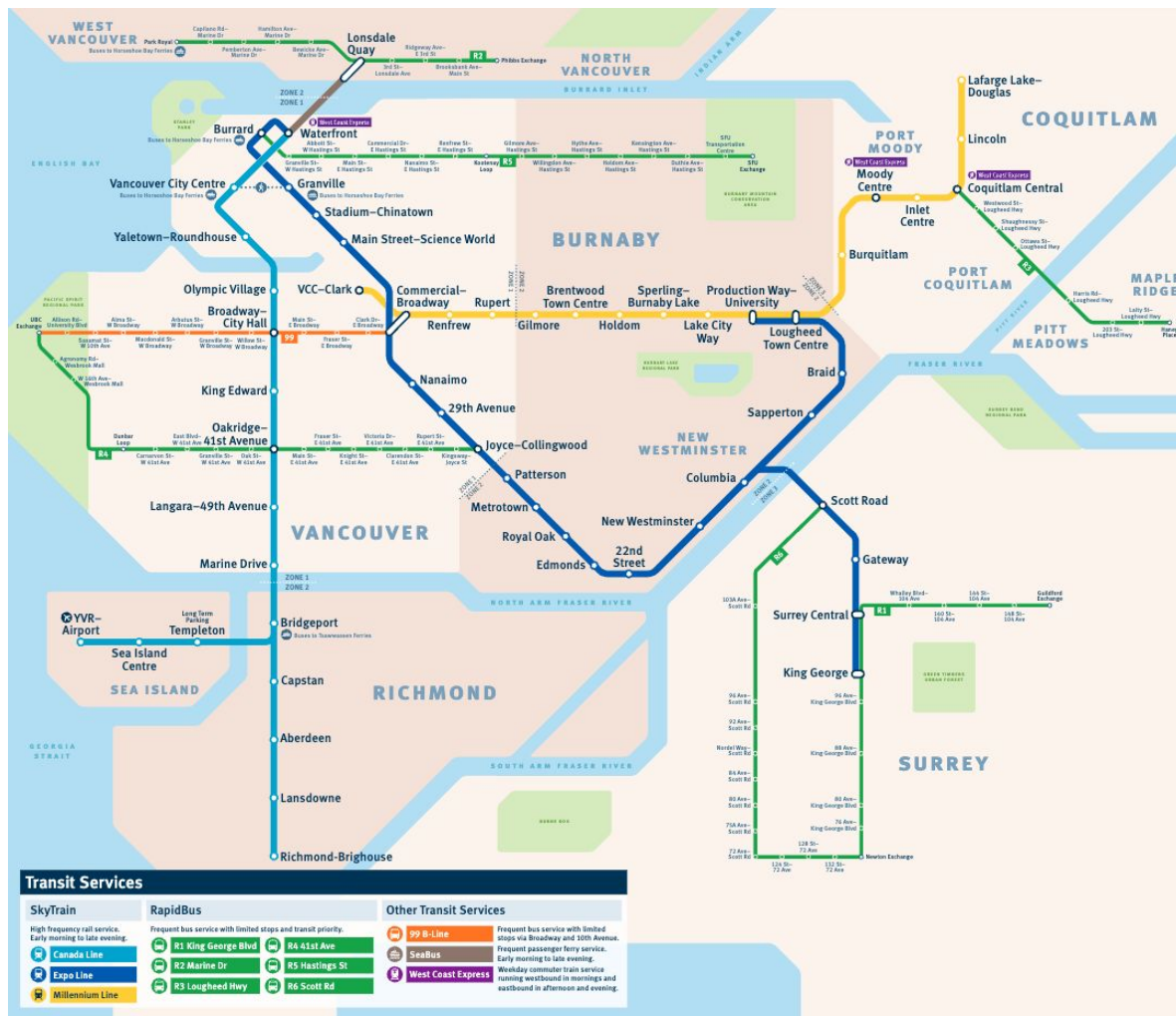
Learning Goals

After this **today's lecture**, you should be able to:

- Define infographics and their role in conveying information effectively.
 - Understand the difference between infographic vs. visualization
- Recognize, define and apply high-level principles of infographic design
- Identify strengths and weaknesses in infographic designs based on high-level principles.

Infographics

Vancouver





What are Infographics?

"Information graphics or infographics are **visual representations of data**, information, or knowledge intended to **present the idea quickly succinctly, and clearly.**" (Wikipedia)

- Static representation that conveys a specific message
- Typically includes graphics and stats (but it doesn't have to)
- Images used do not necessarily have to encode data



Infographics vs. Visualization

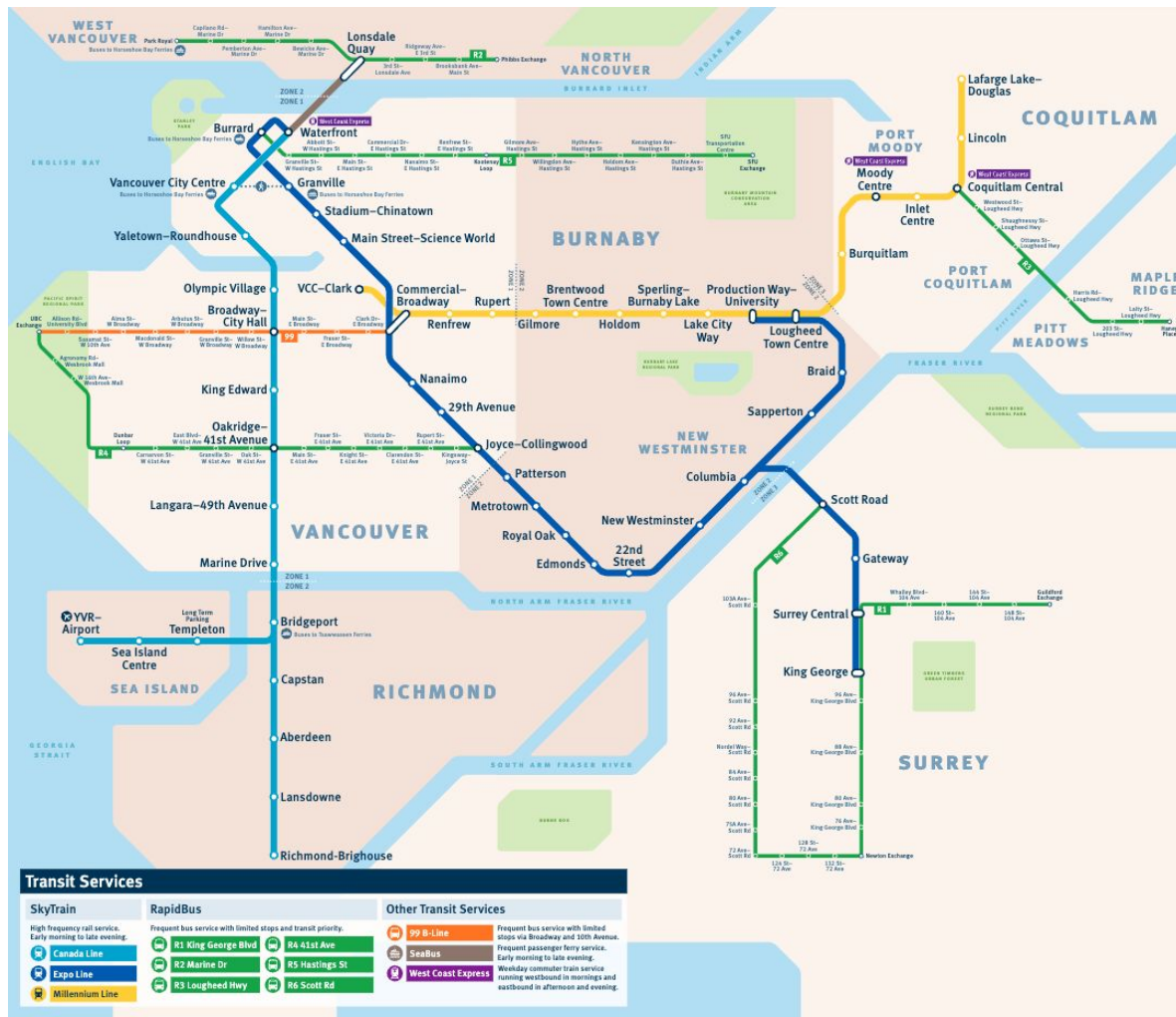
- An infographic is a **static representation** that conveys a specific message. It typically includes graphics and stats but doesn't have to.
- Visualization can be an **interactive or static representation** that primarily uses visual marks to encode data.

Why Infographics?

- Improve Comprehension and Retention
 - mix of text, visuals, and icons to break down complex ideas
- Enhance Engagement
 - attracts more attention than plain text
- Simplify Data and Statistics
 - make numbers more understandable*
 - * = simplification should not compromise the accuracy of the data
- Increase Information Accessibility
 - bridge language and literacy gaps by using icons, symbols, and structured layouts

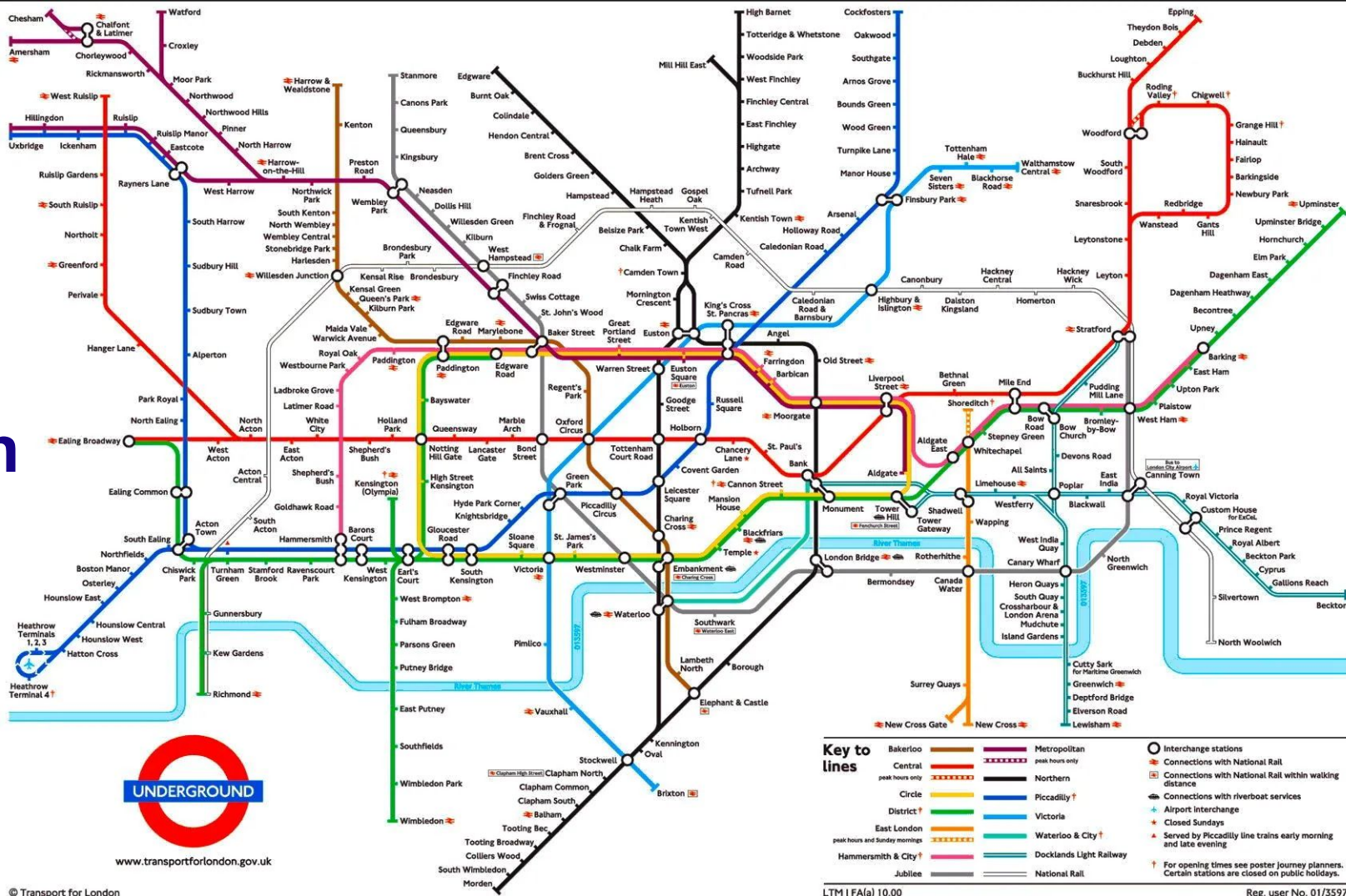
Examples: Transit Maps

Vancouver





London



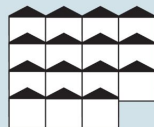
www.transportforlondon.gov.uk



Other Examples

UBC Library Snapshot

UBC Library advances research, learning and teaching excellence by connecting communities within and beyond UBC to the world's knowledge. The Library, a high-ranking member of the Association of Research Libraries (ARL), is the largest library in British Columbia and provides access to expanding digital resources and houses an on-site digitization centre. For more information, visit library.ubc.ca.



15 branches across
2 campuses



315 full-time staff
• 88 librarians
• 184 management & support staff
• 43 student employees

Rankings

- **14 out of 115** university libraries in the Association of Research Libraries (ARL)
- **2nd** among Canadian academic libraries (ARL)
- cIRcle, UBC's information repository, **ranks 2nd in Canada** and 44th globally among 1,650 repositories.

COLLECTIONS

› More than **7.4m** volumes

› More than **1.8m** e-books

› **330,000*** e-journals

› **500,000*** items in

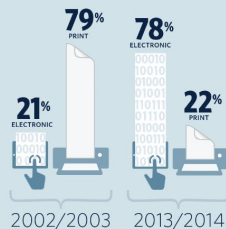
locally produced **digital collections**

More than
5 million e-book downloads
8 million e-journal downloads



LIBRARY RESOURCES

Expenditures shifting
from **print** to **electronic**



on-campus **3.8M⁺**

visits (JAN-DEC 2013)

library.ubc.ca **11.1M⁺**

UBC Library on SOCIAL MEDIA

Twitter **4,8K**

16 accounts

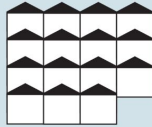
Facebook **9,9K**



Librarians provided
1,768 instructionals
to more than **41,427**
participants
and answered
reference questions
• 54,648 in-person
• 10,756 online

UBC Library Snapshot

UBC Library advances research, learning and teaching excellence by connecting communities within and beyond UBC to the world's knowledge. The Library, a high-ranking member of the Association of Research Libraries (ARL), is the largest library in British Columbia and provides access to expanding digital resources and houses an on-site digitization centre. For more information, visit library.ubc.ca.



15 branches across
2 campuses



315 full-time staff

- 88 librarians
- 184 management & support staff
- 43 student employees

Rankings

- **14 out of 115** university libraries in the Association of Research Libraries (ARL)
- **2nd** among Canadian academic libraries (ARL)
- cIRcle, UBC's information repository, **ranks 2nd in Canada** and 44th globally among 1,650 repositories.

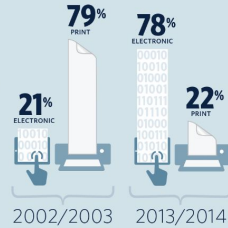
COLLECTIONS

- › More than **7.4m** volumes
- › More than **1.8m** e-books
- › **330,000*** e-journals
- › **500,000*** items in locally produced **digital collections**
- More than **5 million** e-book downloads
- 8 million** e-journal downloads



LIBRARY RESOURCES

Expenditures shifting from **print** to **electronic**



Librarians provided 1,768 instructionals to more than **41,427 participants** and answered reference questions

- 54,648 in-person
- 10,756 online

UBC Library on SOCIAL MEDIA

Twitter **4,8K**

16 accounts

Facebook **9,9K**

Look at all the **amazing things** we did, together ...

4,238,483
individual donations

441,108
volunteer hours logged

57,738
positive actions taken

(recycling, buying products with less packaging, buying groceries for an elderly neighbor, reading a book about inclusion, taking a shorter shower and more!)

10,247,738
total Good Things

1.2+ more than 2020

Company: Benevity

Principles for Infographics



Principles for Infographics

- **Simplicity**
- **Consistency**
- **Visibility**
- **Navigability** (structure)
- **Suitability**



Principles for Infographics

- **Simplicity** – minimal text, clear message, avoid clutter
- **Consistency**
- **Visibility**
- **Navigability** (structure)
- **Suitability**



Principles for Infographics

- **Simplicity** – minimal text, clear message, avoid clutter
- **Consistency** – layout and design elements should be consistent.
 - 2 to 3 font sizes, colour scheme
- **Visibility**
- **Navigability** (structure)
- **Suitability**



Principles for Infographics

- **Simplicity** – minimal text, clear message, avoid clutter
- **Consistency** – layout and design elements should be consistent.
 - 2 to 3 font sizes, colour scheme
- **Visibility** – appropriate font sizes, colors that contrast
- **Navigability** (structure)
- **Suitability**



Principles for Infographics

- **Simplicity** – minimal text, clear message, avoid clutter
- **Consistency** – layout and design elements should be consistent.
 - 2 to 3 font sizes, colour scheme
- **Visibility** – appropriate font sizes, colors that contrast
- **Navigability** (structure) – clear order to follow, use scale/proportion to emphasize key points/headings
- **Suitability**



Principles for Infographics

- **Simplicity** – minimal text, clear message, avoid clutter
- **Consistency** – layout and design elements should be consistent.
 - 2 to 3 font sizes, colour scheme
- **Visibility** – appropriate font sizes, colors that contrast
- **Navigability** (structure) – clear order to follow, use scale/proportion to emphasize key points/headings
- **Suitability** – right data for message, right graphic for the message, right metaphors for the audience



Principles for Infographics

- **Simplicity** – minimal text, clear message, avoid clutter
- **Consistency** – layout and design elements should be consistent.
 - 2 to 3 font sizes, colour scheme
- **Visibility** – appropriate font sizes, colors that contrast
- **Navigability** (structure) – clear order to follow, use scale/proportion to emphasize key points/headings
- **Suitability** – right data for message, right graphic for the message, right metaphors for the audience

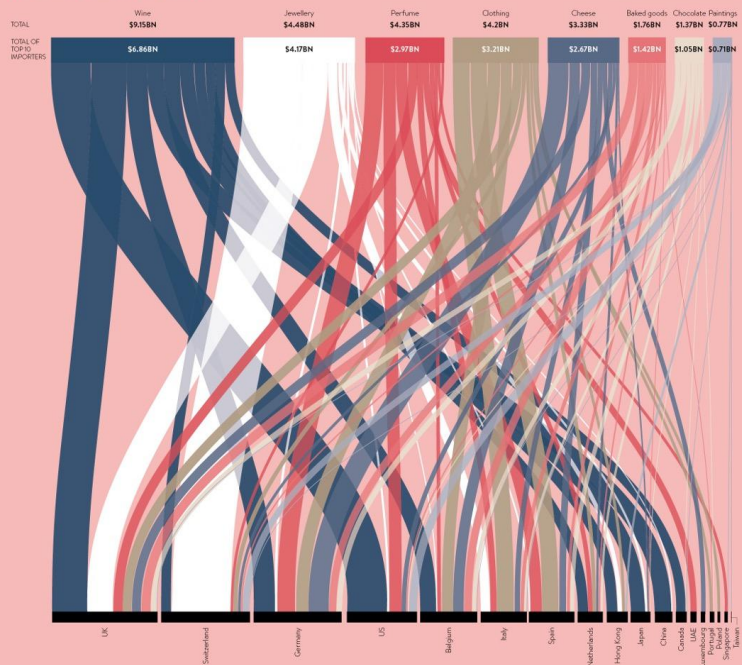
Principles in Action

Made in France

France exported \$572 billion of goods in 2015, making it the sixth largest exporter in the world. While aircraft, cars and medicines are the country's highest-valued goods, there are many other exports to varied markets.

FRENCH EXPORTS AND TOP 10 IMPORTERS

Source: Trade Map/UN Comtrade 2016



ALTERNATIVE EXPORTS AND MOST POPULAR MARKETS

Source: Trade Map 2016

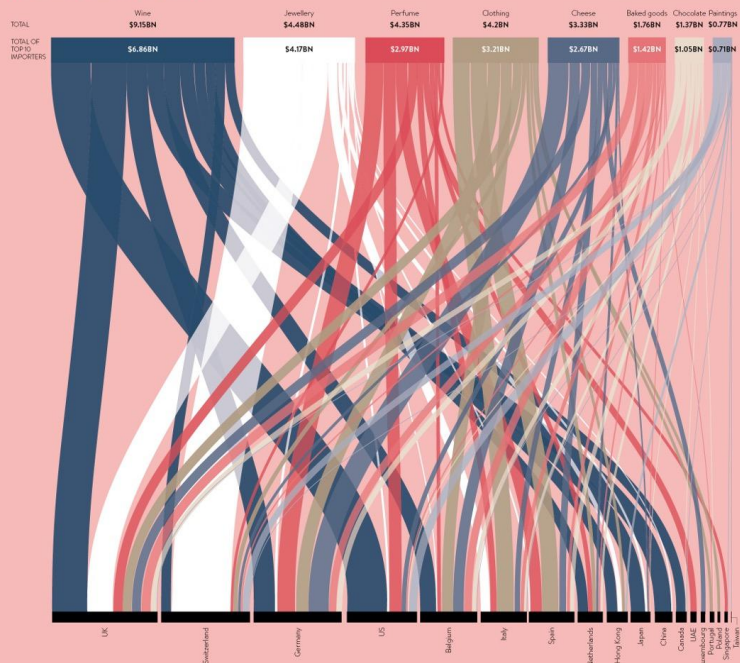


Made in France

France exported \$572 billion of goods in 2015, making it the sixth largest exporter in the world. While aircraft, cars and medicines are the country's highest-valued goods, there are many other exports to varied markets

FRENCH EXPORTS AND TOP 10 IMPORTERS

Source: Trade Map/UN Comtrade 2016



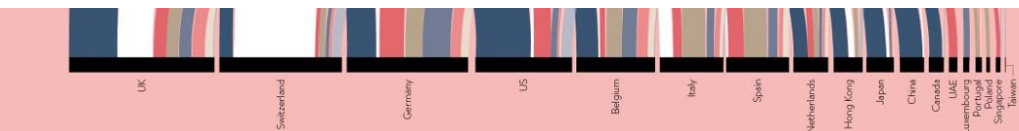
ALTERNATIVE EXPORTS AND MOST POPULAR MARKETS

Source: Trade Map 2016



FRENCH EXPORTS AND TOP 10 IMPORTERS

Source: Trade Map/UN Comtrade 2016



ALTERNATIVE EXPORTS AND MOST POPULAR MARKETS

Source: Trade Map 2016

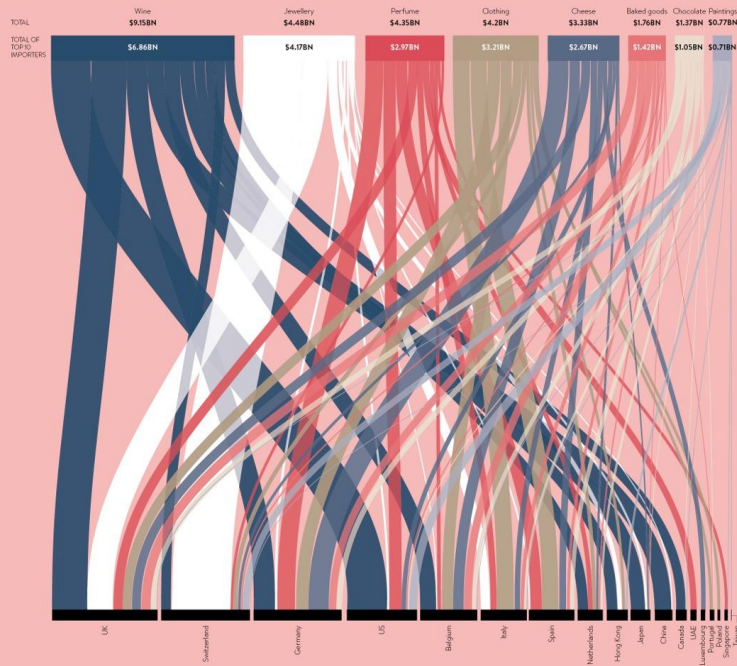


Made in France

France exported \$572 billion of goods in 2015, making it the sixth largest exporter in the world. While aircraft, cars and medicines are the country's highest-valued goods, there are many other exports to varied markets

FRENCH EXPORTS AND TOP 10 IMPORTERS

Source: Trade Map/UN Comtrade 2016



ALTERNATIVE EXPORTS AND MOST POPULAR MARKETS

Source: Trade Map 2016



✗ Simplicity

use of a sankey diagram does not help

✗ Visibility

text at the bottom is very hard to read

✗ Suitability

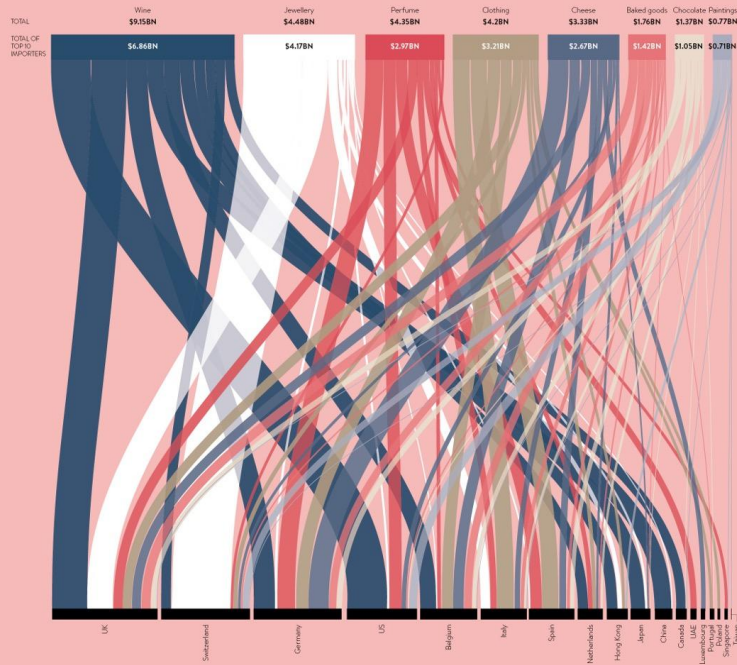
what is the focus of the infographic?

Made in France

France exported \$572 billion of goods in 2015, making it the sixth largest exporter in the world. While aircraft, cars and medicines are the country's highest-valued goods, there are many other exports to varied markets

FRENCH EXPORTS AND TOP 10 IMPORTERS

Source: Trade Map/UN Comtrade 2016



ALTERNATIVE EXPORTS AND MOST POPULAR MARKETS

Source: Trade Map 2016



✗ **Simplicity**

use of a sankey diagram does not help

✗ **Visibility**

text at the bottom is very hard to read

✗ **Suitability**

what is the focus of the infographic?

✓ **Navigability**

top to bottom

✓ **Consistency**

colour theme

Infographics vs. Visualization

- An infographic is not a type of visualization, as it could just be made up of words and images. (e.g., 70% of children like chocolate ice cream with a picture of an ice cream cone).
- The images in an infographic do not necessarily have to encode data. So you can have infographic that has an emoji, an image of an animal, food, but not encode quantitative data with the images (e.g. no bar charts).