

Lesson Plan: Principles of Visualization

Zach del Rosario (zdelrosario@citrine.io)

September 10, 2019

1 Overview

The purpose of this lesson is to help students understand the fundamental principles of visualizing data. This lesson draws on the work of Cleveland and McGill (1985, *Science*) on the *visual hierarchy*, a quantitative ranking of different methods of encoding numeric values in a visual fashion. As a result of working through this lesson, students will be able to:

1. Understand that visualization is accomplished through different *visual encoding methods* – the rungs on the visual hierarchy
2. Understand that different rungs on the hierarchy convey information more or less accurately
3. Use (2) to judge the efficacy of visualizations

2 Outline and Timings

Task	Min
Setup & Introduction	5
<hr/>	
P1: Card sorting	
Q1	5
Q2	2
Q3	5
Survey	2
Wrap	5
	(19)
<hr/>	
P2: Gapminder	
Gapminder A	7
Gapminder B	4
A vs B	5
1 vs 2: walkthrough	1
	(17)
<hr/>	
Wrap	5
(Total)	(46)

3 Resources

- vis-tasks challenge document

- vis-tasks printable sheet
- vis-tasks notes with Gapminder examples
- Cleveland and McGill (1985) Science
- Gapminder data vis

Table for **P1: Card sorting**

- You will iteratively build the following table during Part 1:

1	2	3	CODE	Description
			COM	Position along a common scale
			NON	Position along a non-aligned scale
			LEN	Length
			ANG	Angle
			PIR2	Area
			SAT	Color saturation

4 Part 1: Card sorting

Before: Do the following:

- Prepare to show the following websites (open in tabs):
 - Pages site – Main tool for presenting
 - Cleveland and McGill (1985) – reference for the visual hierarchy
 - Gapminder visualization – to show how the Gapminder project shows their own data
- Hand out cards for pairs

Beginning: Say the following

- "For the next 50 minutes, we are going to work through a guided exercise together. You are going to work together in pairs on a set of questions, and we will share ideas in a full group discussin."
- "The cards you have depict a dataset. These data are from a dataset on imported cars; the 'Count' is the number of vehicles of particular types in the dataset. We are going to focus mainly on how the data are displayed."
- "For the moment, we are going to focus on the same, single variable, displayed in different ways."
- (Show Q1 through Q3 on the Pages site. **Zoom the page for visibility.**)

Q1: (5 Minutes) How is 'Count' displayed?

- Give one answer: "Position along a common scale"
- (While students are working, write the six codes stacked vertically on the board in order, i.e. the CODE column in the Table above.)

– COM, NON, LEN, ANG, PIR2, SAT

- (Walk among students, listen in on their thinking, ask for volunteers to share in the larger group. Aim for 2-3 volunteers.)
- (Bring students back together. Get volunteers to share their ideas.)
- (Fill in the full table to disambiguate codenames:)
 - COM – Common scale
 - NON – Non-aligned scale
 - LEN – Length
 - ANG – Angle
 - PIR2 – Area
 - SAT – Color saturation

Q2: (2 Minutes) Answer the questions:

- "Which is larger? The 'Count' of 'wagons' with 'fwd drive' OR the 'Count' of 'wagons' with 'rwd drive'? (Count of 'fwd wagon' is **larger** than 'rwd wagon')
- "By how much is one larger than the other?" (By about 2-3 counts.)
- (Make sure the Pages site is visible to students; it is difficult to keep these questions in one's head!)
- (Can skip walking among the students; this question is very quick.)

Q3: (5 Minutes) Rank the six visualizations *in terms of how well they help you answer Q2*. Rate the most helpful visualization 1, and the least helpful as 6; do not allow ties between visualizations. **Make sure to rank at least your top 3.**

- (While students are working, fill in the grid lines from rank 1 to 3. Complete the table.)
- (Walk among students, listen in on their thinking, ask why they chose the particular rankings they did.)

Survey: (2 Minutes) How many people Ranked COM 1? Ranked COM 2? Ranked COM 3?, etc. until pattern emerges.

- (Bring students back together. Count hands **for each cell** in the top three rows. There should be a nearly-diagonal pattern.)
- Ask "Why are we seeing this pattern of preference? It has to do with the *visual hierarchy*."

Cleveland and McGill (1985) (5 Minutes)

- "These visualizations are based on the rungs of Cleveland and McGill's *visual hierarchy*. They studied how accurately people could interpret graphs, based on how the data were visually encoded. They arrived at the following hierarchy, in descending order of accuracy:"
 1. Position along a common scale
 2. Position on identical but nonaligned scales

3. Length
4. Angle; Slope (With slope not too close to 0, 90, or 180 degrees)
5. Area
6. Volume; Density; Color saturation
7. Color hue

- Use Pages site to quickly show hierarchy

- (An aside: We asked something slightly different – preference rather than accuracy.)
- publication link
- "We can *use* this insight to help judge and design graphs."

5 Part 2: Gapminder

- Send remote folks to online notes
- "Now, we are going to consider four different variables, displayed on the same graph in different ways.
- "The data are from the Gapminder project, which seeks to educate people about global poverty."

Gapminder A vs B

- Q1: (7 Minutes) Gapminder A
 - (2 Minutes) How are the four variables ‘Population, GDP per Capita, Life Expectancy at Birth, Continent’ encoded visually?
 - * (Have students Think-Pair-Share)
 - (5 Minutes) What observations can you make about the data based on the vis?
 - * (Have students Think-Pair-Share)
- Q2: (4 Minutes) Gapminder B
 - (2 Minutes) How are the four variables ‘Population, GDP per Capita, Life Expectancy at Birth, Continent’ encoded visually?
 - * (Talk through this; ask for volunteers on the spot.)
 - (2 Minutes) What observations can you make about the data based on the vis?
- **A vs B** (5 Minutes)
 - "I find it easier to see the lower life expectancy in ‘Africa’ based on Gapminder B."
 - "With the *visual hierarchy*, we can be more specific than ‘this graph is bad’ – we can note that ‘Continent has fewer levels, therefore it is easier to show with a color scale.’"
 - Same data, same variables, different choice of encoding

Timeseries 1 vs 2 (1 Minute)

- Walkthrough two different visualizations
- "Here the choice is less obvious; depends on which variable I care about more – GDP / capita or Life expectancy."
- **"Visualization is an iterative process."**

6 Finale

- Wrapup (5 Minutes) "What did we talk about?"
- "Variables/numbers are *encoded* visually to make a graph; we choose how to encode"
 - (Gesture to the table to remind)
- "We need to make these decisions when constructing a visualization."
 - (Scroll through the Gapminder examples, use them to illustrate different choices.)
- Lessons:
 - **Use our linear scales preferentially** – reserve for the "most important" continuous variables
 - **Use secondary scales strategically** – use for "secondary" variables, or categorical variables
- **Visualization is an iterative process**
 - We won't get the design perfect the first time.
 - What is "good" depends on our goals; exploratory graphs have very different goals than communication graphs.