Informative versus Persuasive versus Visual Art

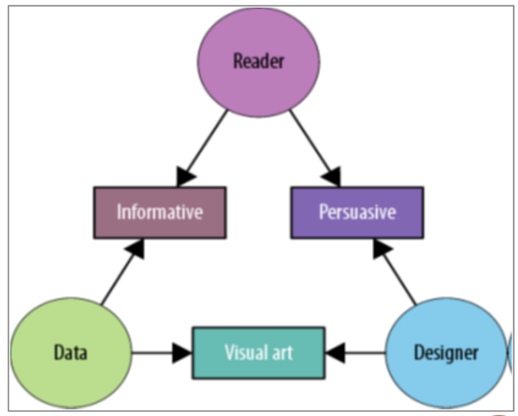
There are three main categories of explanatory visualizations based on the relationships between the three necessary players: the designer, the reader, and the data.

INFORMATIVE VERSUS PERSUASIVE VERSUS VISUAL ART

Categories of explanatory visualizations based on the relationships between the three necessary players:

1. the designer,
2. the reader, and
3. the data.

- Refers to explanatory (or hybrid) visualizations exclusively, because it discusses designing visualizations of data with known parameters and stories.



INFORMATIVE

- An informative visualization primarily serves the relationship between the reader and the data.
- Aims for a neutral presentation of the facts in such a way that will educate the reader
- Informative visualizations are often associated with broad data sets, and seek to distill the content into a manageably consumable form.
- The Burning Man Infographic is an example of informative visualization.

PERSUASIVE

- A persuasive visualization primarily serves the relationship between the designer and the reader.
- Represents a very specific point of view, and advocates a change of opinion or action on the part of the reader.
- The data represented is specifically chosen for the purpose of supporting the designer's point of view, and is presented carefully so as to convince the reader of same.

VISUAL ART

- Primarily serves the relationship between the designer and the data.
- Visual art is unlike the previous two categories in that it often entails unidirectional encoding of information
- Visual art merely translates the data into a visual form.
- The designer may intend only to condense it, translate it into a new medium, or make it beautiful;

Briefly describe the three ingredients for successful visualization.

Designer

Designer is creating visualization *for some reason*. Being aware of motivations, goals, and priorities will help to *design* a successful visualization, rather than merely create an arbitrary visual representation of data. Understanding and defining goal is key to success; Having a well-defined goal will inform subsequent design decisions, and will give you a standard to evaluate your design against. And it will help to make appropriate choices long before the start of assigning axes and plotting points.

Reader

The second source of influence is the reader. As the intended recipient of ideas, the reader holds a very special place in the trinity and can be your biggest ally or your biggest hurdle in clear communication—sometimes both.

It is important to put developer in the shoes of reader, and to take into consideration the unique viewpoint that he will bring with him.

The key questions to ask here are ones like:

- What information does reader need to be successful?
- How much detail does he / she need?
- How long does he / she have to make it effective?

Data

The third source of influence in designing a visualization is your data. The best visualizations will reveal what is interesting about the specific data set you’re working with. Different data may require different approaches, encodings, or techniques to reveal its interesting aspects. While default visualization formats are a great place to start, and may come with the correct design choices pre-selected, sometimes the data will yield new knowledge when a different visualization approach or format is used.

The type of basic questions you will want to ask about your data include:

- Is it a time-series? A hierarchy?

• How many dimensions does it have? Which are the most important ones?

• What sort of relationships do they have (e.g., one-to-one or many-to-many)?

• How variable are they?

• Are the values categorical? Discrete or continuous? Linear or non-linear? How are they bounded?

• How many categories are there?

With an example for each, state the rationale behind the following rules for producing better visualization: Know your audience, Identify your message, Adapt figure to the support medium.

Know your audience: It is important to identify, as early as possible in the design process, the audience and the message the visual is to convey. The graphical design of the visual should be informed by this intent. If you are making a figure for yourself and your direct collaborators, you can possibly skip a number of steps in the design process, because each of you knows what the figure is about. However, if you intend to publish a figure in a scientific journal, you should make sure your figure is correct and conveys all the relevant information to a broader audience.

Identify your message: A figure is meant to express an idea or introduce some facts or a result that would be too long (or nearly impossible) to explain only with words, be it for an article or during a time-limited oral presentation. In this context, it is important to clearly identify the role of the figure, i.e., what is the underlying message and how can a figure best express this message? Once clearly identified, this message will be a strong guide for the design of the figure.

Adapt figure to the support medium: A figure can be displayed on a variety of media, such as a poster, a computer monitor, a projection screen, or a simple sheet of paper (as in a printed article). Each of these media represents different physical sizes for the figure, but more importantly, each of them also implies different ways of viewing and interacting with the figure.

With an example for each, state the rationale behind the following rules for producing better visualization: Do not trust the defaults, Use colors effectively, Avoid Chartjunk.

Do Not Trust the Defaults:

Any plotting library or software comes with a set of default settings. When the end-user does not specify anything, these default settings are used to specify size, font, colors, styles, ticks, markers, etc. Since these settings are to be used for virtually any type of plot, they are not fine-tuned for a specific type of plot. In other words, they are good enough for any plot but they are best for none. All plots require at least some manual tuning of the different settings to better express the message.

Use colors effectively:

Color is an important dimension in human vision and is consequently equally important in the design of a scientific figure. If you decide to use color, you should consider which colors to use and where to use them. For example, to highlight some element of a figure, you can use color for this element while keeping other elements gray or black. This provides an enhancing effect.

Avoid ''Chartjunk''

Chartjunk refers to all the unnecessary or confusing visual elements found in a figure that do not improve the message (in
the best case) or add confusion (in the worst case). For example, chartjunk may include the use of too many colors, too
many labels, gratuitously colored backgrounds, useless grid lines, etc. An element that could be considered chartjunk in one figure can be justified in another. For example, the use of a

background color in a regular plot is generally a bad idea because it does not bring useful information.

Show the contexts in which you would choose the following Visualization techniques: Q-Q Plots, Stacked graphs, Choropleth Maps Give an example for each to justify your response.

Q-Q Plots

The Q-Q plot compares two probability distributions by graphing their quantiles against each other. If the two are similar, the plotted values will lie roughly along the central diagonal. If the two are linearly related, values will again lie along a line, though with varying slope and intercept.

Stacked Graphs.

By stacking area charts on top of each other, we arrive at a visual summation of time-series values—a *stacked graph*. This type of graph (sometimes called a *stream graph*) depicts aggregate patterns and often supports drill-down into a subset of individual series.

A Stacked graph does not support negative numbers and is meaningless for data that should not be summed (temperatures, for example). Moreover, stacking may make it difficult to accurately interpret trends that lie atop other curves. Interactive search and filtering is often used to compensate for this problem.

Choropleth Maps

Data is often collected and aggregated by geographical areas such as states. A standard approach to communicating this data is to use a color encoding of the geographic area, resulting in a c*horopleth map*. One common error is to encode raw data values (such as population) rather than using normalized values to produce a density map. Another issue is that one's perception of the shaded value can also be affected by the underlying area of the geographic region.

Programming Questions

1. **Nov 2017**

Assume that "power.csv" contains information about power production. Fields in csv are "Type of fuel", "Country", "Place", "Mega Watt" and "Cost per Unit". Write the script using pandas, to perform following tasks. List different types of fuels used. Which country generates more power using 'Coal' as fuel used. Which country generates maximum power(including all types of fuels). List the place name which produces maximum power.

```
import numpy as np
from pandas import Series, DataFrame
import pandas as pd
power = pd.read_csv("power.csv")
fuels_used = power['fuel'].unique()
max_coal_power = power['country'][power['fuel'] == 'coal' and power['Mega Watt'] ==
power['Mega Watt'].max()]
place_max_power = power['Place'][power['Mega Watt'] == power['Mega Watt'].max() ]
```

2. **Nov 2018**

Assume that "bill.csv" contains history of sales done in a shop. Fields in csv are "Product Name", "Unit Price", "Quantity" and "Customer Name". Write the script using pandas, to perform following tasks.
Load to DataFrame.
List unique customer name
List Item name with its unit price
List transactions from row number 10 to 20

```
df_bill = pd.DataFrame("bill.csv")
Unique_customers = df_bill ['Customer Name'].unique()
Item_names = df_bill [['Product Name', 'Unit Price']]
Transactions = df_bill [10:21, :]
```

Assume that "Vehicals.csv" contains information about different types of vehicles. Fields in csv are "Type of fuel", "Company", "Model", "Type of Vehicle" and "Cost". Write the script using pandas, to perform following tasks. List different types of vehicles List

companies that manufacture "CAR" as vehicle type Give the costliest "model" among "CAR" List the types of Vehicles company "TATA" produces.

df[‘Type of Vehicle].unique()

df[df[‘Type of Vehicle’]==’car’][‘company’].unique()

df[df[‘Type of Vehicle’]==’car’][‘Cost’].max()

df[df[df[‘Type of Vehicle’]==’car’][‘company’]==’TATA’].unique()

Write python script to perform following actions:

Find indices of non-zero elements from [1,2,0,0,4,0]

Create a random vector of size 30 and find the mean value

Create random vector of size 10 and replace the maximum value by 0

Create a null vector of size 10 but the fifth value which is 1

data = [1,2,0,0,4,0]

print (np.nonzero(data))

z = np.random.random(30)

m = z.mean()

print (m)

z = np.zeros(10)

z[4] = 1

print(z)

Consider the following Python dictionary data and Python list labels:

data = { 'animal': ['cat', 'cat', 'snake', 'dog', 'dog', 'cat', 'snake', 'cat', 'dog', 'dog'],

'age': [2.5, 3, 0.5, np.nan, 5, 2, 4.5, np.nan, 7, 3],

'visits': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],

'priority': ['yes', 'yes', 'no', 'yes', 'no', 'no', 'no', 'yes', 'no', 'no'] }

labels = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j']

Create a DataFrame df from this dictionary data which has the index labels.

Change the age in row 'f' to 1.5.

Select the rows the age is between 2 and 4 (inclusive)

Count the number of each type of animal in df.

```
df = pd.DataFrame(data, index=labels)
df.loc['f', 'age'] = 1.5
df[df['age'].between(2, 4)]
df['animal'].value_counts()
```