Data Abstraction and Color

Q1: What type of attribute (categorical, ordinal, quantitative) are the following? If the attribute is ordinal or quantitative, what is the ordering direction (sequential, diverging, cyclic)? What color map (#1~#6) is the best choice to encode each attribute? (30%)

1. Academic letter grade (e.g. A+, A, A-, B+ B …)

Ordinal data, Sequential , 3

1. Media companies (e.g. Netflix, Walt Disney, Fox, CBS, …)

Categorical data, 4

1. Acceleration (e.g. +2 m/s2  , -3 m/s2 , +22 m/s2 )

Quantitative data, Sequential , 3

1. Scores of an exam (e.g. 90 points, 60 points, 30 points)

Quantitative data, Sequential , 3

1. Survey options (e.g., strongly agree, agree, disagree, strongly, disagree )

Ordinal data, Sequential , 3



#1



#2



#3



#4



#5



#6

Q2: Check this website about mental health survey: (36%)

[https://www.kaggle.com/datasets/mahirahmzh/starbucks-customer-retentionmalaysia-survey](https://www.kaggle.com/datasets/mahirahmzh/starbucks-customer-retention-malaysia-survey)

* Write down the attribute types of the following attributes (categorical, ordinal, quantitative) and explain your choice for every attributes.
  + Id categorial, since the Id is purely used for identification purposes
  + Gender categorical, since genders have different categories.
  + Age ordinal, since the data is separated into discrete groups and has a order of a ranking
  + Status categorical, the data is divided into different categories and the number doesn’t denote a rank or order
  + Income ordinal, the data separate different income levels into orders
  + visitNo ordinal, since the data has a specific order of visiting frequency
  + method categorical, while the data has a numerical value(0,1,2,3,4), the data doesn’t represent a specific order
  + timeSpend ordinal, the data is separated into different intervals of time spend
  + location ordinal, the data represents different levels of distance from the shop
  + membershipCard categorical, since the data only has two categories
  + spendPurchase ordinal, the data has different ranges of the spend
  + produceRate ordinal, since the data is separated into different levels of rate

Visual Channel and Mark:

Q3: You will identify data items and the marks used to encode them, and data attributes and the channels used to encode them. (34%)

For each chart, fill in

* Visual channels used?

◼ Channel X encodes attribute Y ◼ Channel X encodes attribute Y ◼ ....

* Marks used?
  + - Mark of type X encodes item Y
    - Mark of type X encodes item Y
    - …

Note that underneath each chart there is a link to a web page providing details about the project, including information about the data and in many cases supporting interactivity. For now, just encode what you see in the static image in this document; you might notice that interactivity is being used a lot for emphasis (hover highlight, animated transitions, reordering, details popups, small multiples zooming). Think about how you would reconstruct the dataset underlying this picture. Assume one mark per item, think of items as rows of a table (or nodes in a network). Assume each channel is showing an attribute, where channel is a column in a simple data table or network. Remember that multiple channels could redundantly encode the same attribute.

A.

一張含有 螢幕擷取畫面, 文字 的圖片

自動產生的描述

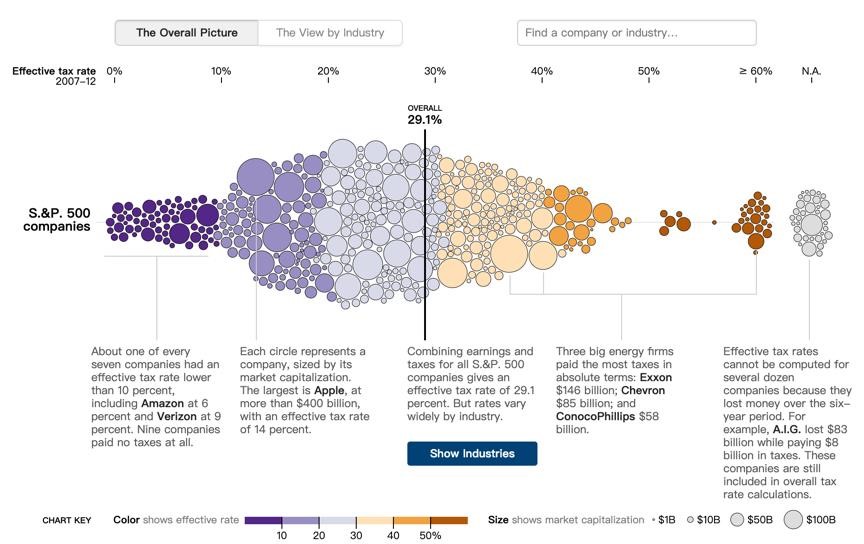
**Visual Channels Used:**

* The **length** of each bar encodes the duration of each storm.
* The **position along the y-axis** of each bar encodes the number of storms in each year.
* The **color** of each bar encodes the type of storm. Different colors represent different types of storms (blue for tropical storms, orange for hurricanes, and red for major hurricanes).

**Marks Used:**

* The **horizontal bars** represent each storm. Each bar is a mark that encodes a storm.
* The **colors** of the bars encode the type of storm (tropical storms, hurricanes, major hurricanes).

B: [https://archive.nytimes.com/www.nytimes.com/interactive/2013/05/25/sundayreview/corporate-taxes.html](https://archive.nytimes.com/www.nytimes.com/interactive/2013/05/25/sunday-review/corporate-taxes.html)



**Visual Channels Used:**

* The **position along the x-axis** of each circle encodes the market capitalization of each company.
* The **position along the y-axis** of each circle encodes the effective tax rate of each company.
* The **color** of each circle encodes the effective tax rate of each company. Different colors represent different tax rates.

**Marks Used:**

* The **circles** represent each company. Each circle is a mark that encodes a company.
* The **colors** of the circles encode the effective tax rates for each company.

C:<https://projects.fivethirtyeight.com/2020-election-forecast/>

一張含有 文字, 螢幕擷取畫面 的圖片

自動產生的描述

Quiz credit: Tamara Munzner

**Visual Channels Used:**

* The **length** of each path segment encodes the margin of victory in each state. Longer segments represent states with a bigger margin.
* The **position** of each path segment along the path encodes the order of states based on their forecasted votes. States with a bigger margin are at the ends of the path, while tighter races are in the middle.
* The **color** of each path segment encodes which candidate is forecasted to win in each state. Different colors represent different candidates.

**Marks Used:**

* The **path segments** represent each state. Each segment of the path is a mark that encodes a state.
* The **colors** of the path segments encode the forecasted votes for each state.