Interactive Data Visualization Report

Course: Interactive Data Visualization (Spring 2020)

Student: Diana Crowe (student nr.: 012056152)

Interactive Visualization found at:
https://public.tableau.com/profile/diana.crowe#!/vizhome/SanFranciscoPoliceIncidents/SanFranciscoPoliceIncidents

1. The chosen dataset, and data abstraction;

I chose a dataset found from Kaggle: San Francisco Crime Dataset
https://www.kaggle.com/roshansharma/sanfranciso-crime-dataset

The csv file contains the information for 2016 about the Crime rates in different regions (Police Districts) of San Francisco with some other important aspects related to crime in a total of 13 columns.

Reasons for choosing: the dataset is very complete (no blank cells), it is well documented/ annotated, and really interesting.

I did some exploratory data analysis and verified that no data cleaning was needed for my chosen dataset. It is also already well annotated and belongs to an open database so there are no issues with licenses.

Description of the columns:

IncidntNum	Incident Number
Category	Category of Crime
Descript	Description of Crime
DayOfWeek	Day of Week when the crime happened
Date	Date
Time	Time
PdDistrict	District
Resolution	Kind of Punishment given to the criminal to resolve the case.
Address	Address where the crime scene happened.
X	Latitude of the crime Location
Y	Longitude of the Crime Location
Location	Exact Location Name
PdId	Pd Id

Data Abstraction is the reduction of a particular body of data to a simplified representation of the whole. Abstraction, in general, is the process of taking away or removing characteristics from something in order to reduce it to a set of essential characteristics.

Deciding what data to use:

- I needed a map frame for my data. I found a shapefile for the San Francisco Police Districts that fit with the year of the data (2016) at:

  (https://data.sfgov.org/Public-Safety/Current-Police-Districts/wkhw-cjsf)

- Out of the columns of data available I used only used "Time" and "PdDistrict". The number of incidents per police district came from (automatically) counting how many times the name of the Police District was listed

Next came adding my data to the shapefile and connecting the two by Police District (Tableau).

2) Design rationale of the visual representation and interaction

Here are the details of the visual representation and interaction that I settled on:

- The data visualization consists of two parts, which are linked:
  - **Part 1:** Map of San Francisco
    - the borders of all the different Police Districts are indicated (polygons).
    - each district is colour-coded to indicate crime/incident rates
    - when you hover the mouse over a district you get a floating pop-up rectangle with extra information on the name of the district and number of incidents
    - colours: orange-blue divergent reversed, with full colour range. It looks nice, shouldn't pose problems to colour-blind people, the colours have a good contrast so that one can notice differences easily, and it is also a very intuitive colour palette to use. The blues are the lowest amounts of crime and the oranges/reds are the highest amounts of crime, which is reminiscent of temperatures: blues for colder (lower) temperatures and red for hotter (higher) temperatures. Area of the map outside San Francisco is displayed in muted colour to create negative space and draw attention to the city areas and the data that we want displayed.
  - Part 2: Chart of Incidents throughout the day
    - Displayed the number of incidents at each hour of the day, per Police District.
    - Matching colour scheme to the map: orange-blue divergent reversed, with full colour range. It is less confusing to have only one colour scheme, and it also ties the two displays together.
    - No grids on the background and just plain white background so it doesn't distract the viewer.

        - When you hover the mouse over the chart it displays a floating rectangle with information on which district it corresponds to, the hour of the day, and the number of incidents

The two parts of the visualization are linked and work as filters for each other:

- **Map**: when you click on a police district on the map it gets highlighted. It means that all other police districts get muted out on the map and, at the same time, the chart below also highlights that police district /mutes out the display for the other police districts. Click in an empty are of the map to reset.
- **Chart**: when you click on one of the lines of the chart, it selects just the line of that district in the chart (all other lines get muted out) and also selects only that district in the map above so that all other districts get muted out.

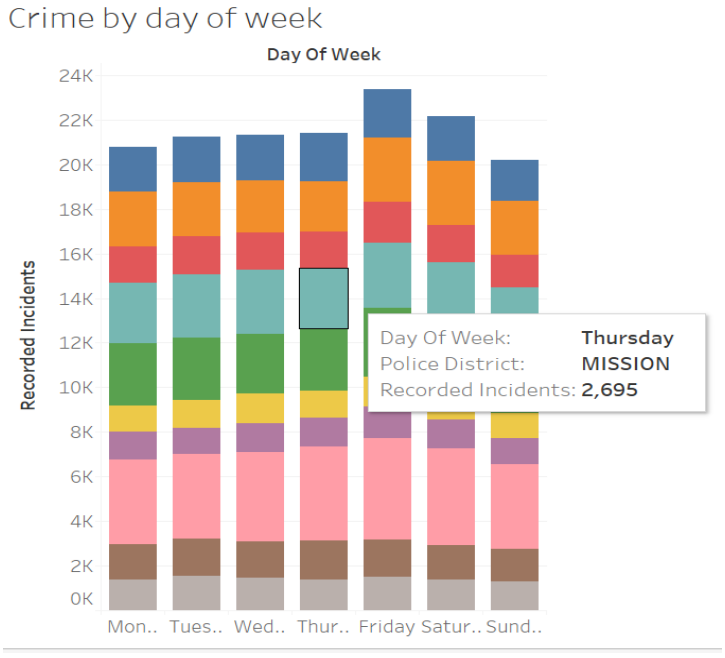
  If you click on one of the hours of the day in the chart, it displays only the data for that hour (all districts) in the chart and it will also change the map above to display (for all districts) only the incidents for that hour of the day.

  Click in an empty area of the chart to reset.

3) Iterative process

- Changes/iterations made to my visualization project:
    - Created an interactive map visualization in Tableau where I had all the Police Districts of San Francisco colour-coded for amount of incidents and, if you move the mouse over each district, you see the information for the name of the Police District and for the number of records for that district.
    - Renamed “Number of records” to “Recorded Incidents”
    - Renamed “District” to “Police District”
    - Created a second sheet in Tableau and named it “Crime throughout the day”. I have “SUM(Recorded Incidents)” in the rows, “HOUR(time)” in the columns and also dragged “PD District” onto the canvas to differentiate the fluctuation of incidents throughout the day for each separate police district.
        - Later renamed the sheet “Incidents throughout the day” since not all incidents that require police are criminal, or if they potentially are criminal, that should still be left for the courts to decide. With that same rationale, I renamed the first tab “Police Incidents in San Francisco” instead of “Crime in San Francisco”.
    - Combined the two sheets (sheet 1 with the map of San Francisco displaying number of incidents per police district, and sheet 2 with the variation of the number of incidents depending on the time of the day) into a dashboard
    - Linked the two displays to make just the one visualization by making the map with the police districts into a filter for the second elements (the chart with the number of incidents in each district depending on the time of the day)

- I wanted the colour schemes in the two linked graphics to match, so I started by applying the same colour scheme to the chart as I had in the map (red-black-white diverging, reversed).
    - Unfortunately, since now I had sections of the graphic that were displayed in white over a white background, this was not a good choice for a colour scheme
    - Changed both the map and the chart into a temperature diverging colour scheme. It looked really nice.
    - I then worried whether there might be colour-blind people who would not be able to differentiate between the red and green of my temperature display (which varies from green to red)…
    - Changed the colour scheme of both charts to orange-blue divergent reversed, with full colour range. It looks nice, shouldn't pose problems to colour-blind people, the colours have a good contrast so that one can notice differences easily, and it is also a very intuitive colour palette to use. The blues are the lowest amounts of crime and the oranges/reds are the highest amounts of crime, which is reminiscent of temperatures: blues for colder (lower) temperatures and red for hotter (higher) temperatures.
- Experimented with adding extra information. I charted the number of incidents per day of the week, differentiated by police district. Tried a few ways of incorporating this with the previously charted and linked data, but it looks very convoluted and confusing to have three displays as part of the same visualization. I also couldn't get a matching colour scheme.



- Found out that I could link the two graphics the other way around – aka, make each graphic a filter for the other. If you click on a district on the map with the police districts, it filters just that district for the chart with crime throughout the day. If you click on a time of the day in the chart, it filters that in the map and displays in the map just the incidents for that time of the day.

Diana Crowe (student nr.: 012056152)

4) Lessons learned

General:

- 65% of the population are visual learners.
- The average attention span of a person is 8 seconds so it's important to get our message across within that allotted time. Visualizations convey the essential message much faster than text, but they should be as simple as possible while still being visually stimulating.
- Less is more. Keep it simple, else you risk distracting from what you want to convey. Present only the most useful, relevant information in the clearest way possible.
- Keep in mind what kind of questions a user might be wanting to answer with the visualization.
- The design needs to have all the visual elements balanced: shape, colour, negative space and texture equally distributed.
- Key areas must be emphasized to draw the user's attention to the right data points (in general, place most important data in top left corner as a user's gaze falls there first).
- Pay attention to details: size, colours, graphics, textures, placement, etc -to direct the user to the most important pieces of information.
- Emphasize similarities through the use of patterns (such as similar colours). This also draws attentions to anomalies since they will be a break in the pattern.
- When possible, include short narratives to emphasize the key insights and provide context for interpretation.
- Interactivity should help to clarify, not distract or confuse the data presentation.

Tableau:

- One may need to find/use extra datafiles (such as shapefiles) in order to properly display geographical data, even if the latitudes and longitudes were included in the data
- In current versions of Tableau, you can only connect to point geometries, linear geometries, or polygons. You cannot connect to mixed geometry types.
- You can link/connect/filter multiple elements to create one unified visualization.

Text:

- Use useful, clear, explanatory titles to help understand the intention of the data presentation and focus on the right questions.
- Most people read from top left to right and gradually down the page, so in static visualizations it is good to mimic movement in an F pattern, which can also be done by the use of colours.
- Use short and descriptive titles on the upper left corner.
- All text should be horizontal, which includes titles, subtitles, annotations, and data labels. Exceptions can be made for line labels and axis labels. One should consider switching graph orientation (e.g., from column to bar chart) to make text horizontal.

- The text size should be hierarchical and readable. Subtitles and annotations (call-outs) can add extra information for explaining and helping interpret the data or highlight data points.
- The text should sufficiently contrast the background.
- The displayed data should sufficiently contrast the background.

Graphs:

- Graphs should be free from clipart, fancy background colours or other illustrations used solely for decoration.
- Gridlines should be avoided and, if present, are muted (I didn't use any)
- Axes should not have unnecessary tick marks or axis lines
- The data should be displayed in an order that makes logical sense to the viewer.
- Graphs should start at zero if possible and the spaces between axis intervals should be in the same units, even if not every axis interval is labelled.
- The type of graph should be appropriate for the data. For example, line graphs for data over time.
- The individual chart elements should work together to reinforce the overarching takeaway message. Choices about graph type, text, arrangement, colour, and lines should reinforce the same takeaway message. – This is why I made sure that the two elements of my Visualization had the same colour scheme and were linked.

Colours:

- Some colour schemes work great for some visualizations but not for others. Revisit colour-schemes after choosing to make sure that: there is enough contrast so that it is easy to distinguish the nuances, the colours won't cause problems for colour-blind people, there are not so many colours so as to make the whole display confusing, etc.
- Colour should be legible for people with colour blindness.
- Action colours should guide the viewer to key parts of the display. Less important, supporting, or comparison data should be a muted colour, like grey.
- Criminals like to sleep too. For all police districts the lowest number of incidents is between 1am and 7am.

5) References

- Data set from Kaggle - Dataset https://www.kaggle.com/roshansharma/sanfranciso-crime-dataset
- What is data abstraction? - https://whatis.techtarget.com/definition/data-abstraction
- Source for San Francisco Police Districts shapefile - https://data.sfgov.org/Public-Safety/Current-Police-Districts/wkhw-cjsf
- Data Visualization checklist - https://stephanieevergreen.com/updated-data-visualization-checklist/
- Lecture material from the course "Interactive Data Visualization"
- Tableau training: https://www.tableau.com/learn/training/20201 and many others