

Smoke & Alcohol Visualization

Information Visualization
20201233 이정목(JungMok Lee)

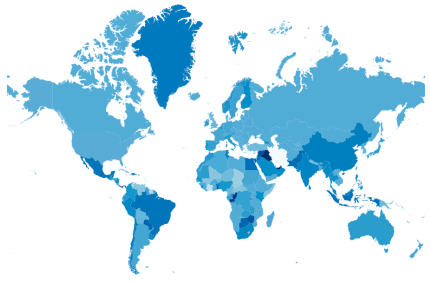
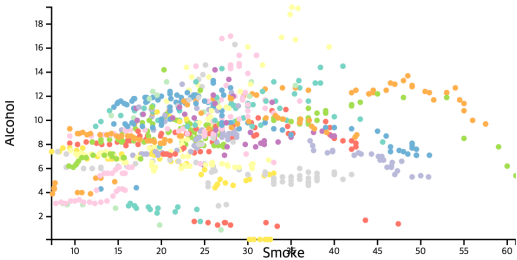
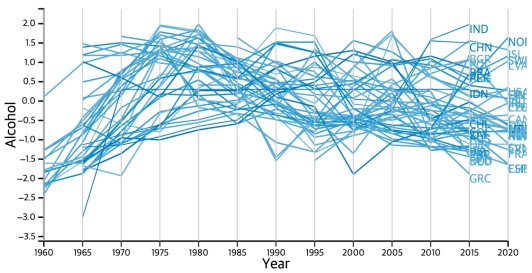
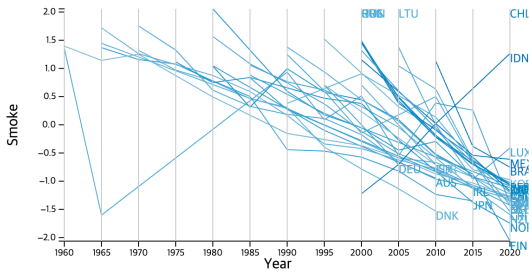
Smoke & Alcohol Visualization

Parallel Coordinate Chart & ScatterPlot & Choropleth Chart Bubble & Choropleth Play

☒ Select All

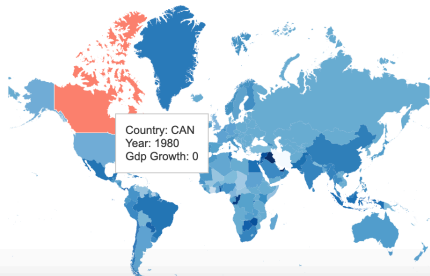
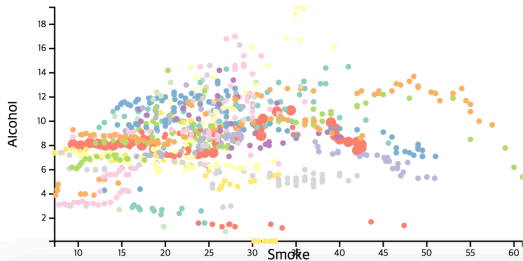
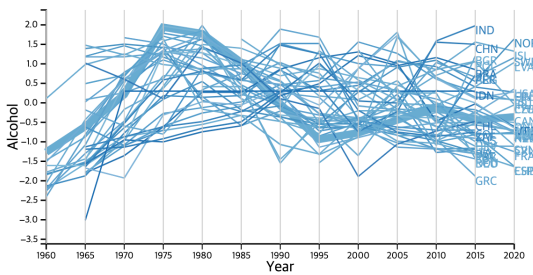
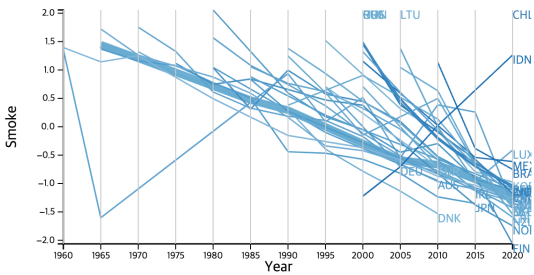
1960 Year: 1980

2020



1960 Year: 1980

2020



I was quite interested in how smoke and alcohol consumption changes along the year, so I wanted to show this using what I learned during information visualization. I have made the GitHub pages for this visualization. You can check here : <https://mok0102.github.io/infovis-smoke-alcohol/>

Data

For this, I used these datasets : smoke consumption, alcohol consumption, population, gdp growth and geojson. Since I wanted to show how smoke & alcohol consumptions along the year and give an association of this trend with the gap growth, I chose those dataset. The population is for drawing scatterplot, the radius encoding in animation is done with population. The geojson should have the ISO_A3, since my other dataset only has ISO_A3 as a country name, so I researched on the google for this geojson.

Smoke consumption : <https://data.oecd.org/healthrisk/daily-smokers.htm#indicator-chart>

Alcohol consumption : <https://data.oecd.org/healthrisk/alcohol-consumption.htm#indicator-chart>

Population : <https://data.oecd.org/pop/population.htm>

GDP growth: <https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=OE>

Geojson : <https://github.com/datasets/geo-countries/blob/master/data/countries.geojson>

For drawing the scatterplot, since I didn't want to give the whole three smoke, alcohol, population in drawBubble, I combined all three using python. (Since I am more familiar with python, I just made the new csv file using below code). Others, I preprocessed dataset such as changing them into integer, float(using +), and year string to js time. And I filtered out when the year's remainder is not 5 for smoke and alcohol parallel coordinate chart.

```
import pandas as pd

# Read the CSV files
smoke_df = pd.read_csv('smoke.csv')
alcohol_df = pd.read_csv('alcohol.csv')
population_df = pd.read_csv('population.csv')

# Merge the dataframes based on location and year
merged_df = pd.merge(smoke_df, alcohol_df, on=['LOCATION', 'TIME'])
merged_df = pd.merge(merged_df, population_df, on=['LOCATION', 'TIME'])

merged_df = merged_df.rename(columns={
    'LOCATION': 'Country',
    'INDICATOR_x': 'Smokers_Indicator',
    'SUBJECT_x': 'Smokers_Subject',
    'MEASURE_x': 'Smokers_Measure',
    'FREQUENCY_x': 'Smokers_Frequency',
    'Value_x': 'Smokers_Value',
    'Flag Codes_x': 'Smokers_Flag_Codes',
    'INDICATOR_y': 'Alcohol_Indicator',
    'SUBJECT_y': 'Alcohol_Subject',
    'MEASURE_y': 'Alcohol_Measure',
    'FREQUENCY_y': 'Alcohol_Frequency',
    'Value_y': 'Alcohol_Value',
    'Flag Codes_y': 'Alcohol_Flag_Codes',
    'INDICATOR': 'Population_Indicator',
    'SUBJECT': 'Population_Subject',
    'MEASURE': 'Population_Measure',
    'FREQUENCY': 'Population_Frequency',
    'Value': 'Population_Value',
    'Flag Codes': 'Population_Flag_Codes'
})

merged_df.to_csv('smoke_alcohol_pop_merged.csv', index=False)
```

My application

My application is about showing the smoke & alcohol consumption along the year.

For those, you can select the year from slider. If you select the year from the slider, corresponding scatterplot and choropleth map is updated. Since two parallel coordinate chart shows overall years, it should not change.

Also you can select some countries if you want. I added "select all" button so if the user want to select/unselect all, it is done more easily. And you can see only selected countries. As you add/subtract some countries from checkbox, the parallel coordinate charts, scatterplot, choropleth map changes correspondingly.

There are two buttons. First button shows the current layout's charts. Second button is for animation of scatterplot & choropleth. It is used to see the scatterplot & choropleth animated trends.

Now, I will explain about chart.

The first parallel coordinate(one the left) shows the smoke consumption along the year for all selected country. It shows the tendency of decreasing so I used gdp on the color channel. And it is normalized according to its year.

Similarly, the parallel coordinate chart of alcohol consumption along the year for selected country is also shown on the right. If the user hovers on certain line, the line is thickened and becomes easier to see tendency of the line, and bidirectly linked to the smoke parallel coordinate chart & scatterplot & choropleth map. So you can easily see how the trend changes along the year using interaction.

Next is the the scatterplot which shows the smoke consumption on x axis and alcohol consumption on y axis for certain selected year. I added some animation to highlight circle along the year in scatterplot & Choropleth Play button.

Next is the choropleth map. The reason why I chose choropleth map is that is will be very interesting interactive if the user can hover on certain country in choropleth map and the corresponding line and circle is highlighted. Also my choropleth map shows curYear's gdp growth. So you can check what it looks like in other graph if the gdp is high/low.

Visual Encoding

The reason why I chose the parallel coordinate charts for first is that I wanted to see the overall trends for smoke and alcohol consumption. Since two parallel coordinate charts show different information and the lines are too much clustered, I wanted to give highlight & two parallel coordinate charts work interactively. So if user hovers on certain parallel coordinate chart's line, the corresponding country's line in the other chart is also highlighted. For this, I made the global variable smokeSvg, and alcoholSvg, and selected each line through filter and gave thick line-stroke(8).

Since the line chart is not so visible, I added normalization and used parallel coordinate chart. The un-normalized chart did not show some trend, the normalized line shows some tendency through the cluster.

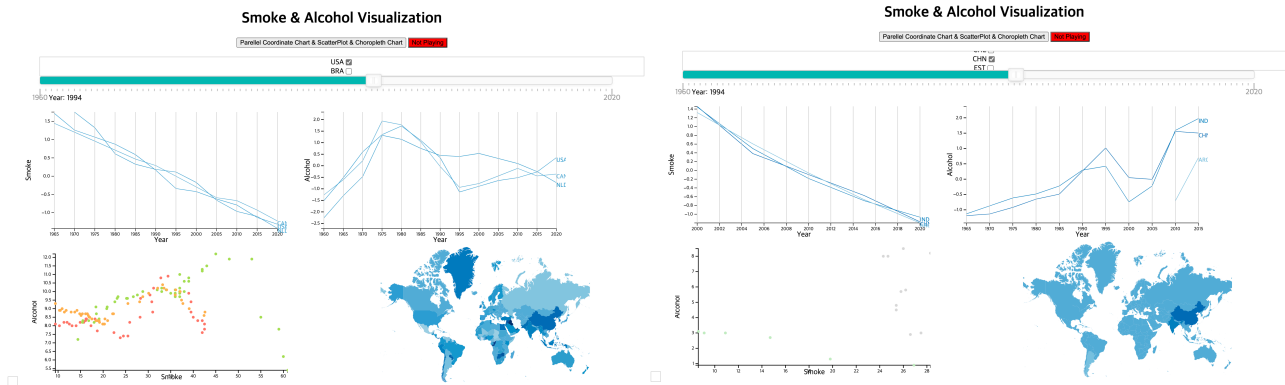
And the scatterplot is to show relation between smoke and alcohol consumption and the animation transition over the year. So if you click the play button after looking at the normalized parallel coordinate chart, the scatterplot's circles are highlighted along the year and shows the left-moving tendency, which I intended.

Next is choropleth map. The reason why I selected the choropleth map is that any chart without country's position encoding seems not so effective when the user want to find wanted country's tendency. The user may not know which country's are wanted by himself/herself(so it can be explore or browse), so the choropleth map is added to help user explore/browse through the map. So when the user hovers on certain country, the tooltip is shown and also the corresponding parallel coordinate chart & scatterplot are highlighted, which makes user to analyze more easier.

Scenarios & Observation

First I want the user to experience the whole information such as parallel coordinate charts for alcohol and smoke consumption over the years, and scatterplot 1980, and the gdp growth map as choropleth. In here, the user will first notice that the smoke consumption decreases over the year, while alcohol consumption is not. And I want the user to look each of them precisely, using the interaction. When the year in slider is set to 1980(which is default), I want the user to hover the mouse on the some circles on scatterplot. Then user can see some clusters in parallel coordinate chart, especially alcohol parallel coordinate chart. Then user can easily find out smoke consumption just decreases, and alcohol consumption slightly increases in 1975-1985 for most countries. As you hover on scatterplot's USA(orange), and NLD(right green) you can see the trend interactively. For now, I want to the user to move the year slider to 2000, then user can see the choropleth map has changed. If the user hovers on the china(or other dark countries such as India, russia) in choropleth, you can see the china has the different tendency that the alcohol consumption increases on 2000s, not in 1980s. So you can find the trends that there are two types for countries: early alcohol consumption and later alcohol consumption. The early alcohol consumption countries show the alcohol consumption tendency of increasing in 1980,

and decreases and remain in 2000s. The later alcohol consumption countries show the alcohol consumption tendency of increasing in 2000s. Which is quite interesting. The one feature of later alcohol consumption countries is that the gdp growth starts rapidly increasing over 1980-2010, such as ARG(Argentina), CHN(China), IND(India). So to conclude, while the smoke consumption is just decreasing, the alcohol consumption has shown that two type of tendencies, which are related to the gdp growth.



First: early alcohol consumption country
 Second : later alcohol consumption country

And user may want to click the Bubble & Choropleth Play button. Since the two parallel coordinate charts are showing the whole year's tendency, it does not change. But the scatterplot and choropleth changes as the slider moves automatically. The tendency that can be seen in scatterplot is that the circles are starting to move left(which means the overall smoke consumption for countries are decreasing), while the alcohol consumption has not shown the dynamic change. So I tried to show the smoke consumption decreasing tendency in two way : parallel coordinate chart(smoke consumption), and scatterplot's animation. I think both way are effective. And looking at the choropleth map, you can see that some countries so a dark color for most of the years, such as china and India, and argentina. As I said earlier, you can find tendency through the related interaction if you hover on those countries, you can see those countries are mostly later alcohol consumption countries.

Lastly, user can unselect select all and select only wanted countries, as shown above. User can see precisely and interactively hover on parallel coordinate chart & bubble & map, and all four graphs interacts.

Requirements

This visualization is developed in m1 Mac Pro (16:10). The layout is maintained in the full screen, so please try this in full screen instead of half / or smaller screen.