Milestone Three

Week Nine - Project Three

Ross L. Kim-Schreck

DSC680 Applied Data Science

Professor Iranitalab

2024.08.10

Exploring ways to curb the suicide rate in the Republic of Korea

Business Problem

In the Republic of Korea, which I will refer to as 'the ROK', the suicide rate has been increasing throughout the ROK's modern history and is now the highest in the world. The country also ranks second in declining birth rates. The ROK is among the highest in the growing elderly population as well. With all these factors, the work force is steadily dwindling, potentially having catastrophic effects on the economy. The ROK's population is expected to rapidly decline from 2023 onwards. In the long run, this will have massive implications on the demographics of the country. With a declining population, suicide must be prevented for economic and social reasons. The issue not only negatively affects persons related, but also due to the number of suicides, affects society as a whole, businesses, the healthcare industry, and the economy.

Background / History

The Republic of Korea was a war-torn third-world country in the mid 20th century. After the partition of the Korean peninsula post World War Two, the two emergent Koreas were stark contrasts of each other. The North was developed by the previously-occupying Empire of Japan, and had been using the North as an industrial region to produce arms for their war machine. The south was used for its more fertile soil as farmland to feed its growing population. Around this time up until roughly the 1990s, Japan also

had the highest suicide rate. The spirit of the kamikaze became prevalent in the common citizen of Japan. Since the Korean war, the ROK and DPRK (North Korea) were initially both military dictatorships despite the ROK being officially a democracy. It wasn't until the military dictator Park Jeonghui decided to take an economy-first initiative. The ROK, through its rapid industrialization, experienced an economic boom throughout the 1970s. While undergoing rapid economic development as a military dictatorship, the ROK began organizing its population and infrastructure largely based on the infrastructure that Japan had already provided as a foundation. The legal and educational systems were modeled on the Japanese systems. Korea's cultural foundation of confucianism combined with the strict and rigid educational and judicial systems provided by Japan began causing stress on its population. East Asia's already competition-heavy education system along with ultra conservatism in society led many to unrealistic expectations in academics and professional life. To this day, despite the ROK having become a fully fledged democracy, Koreans do not benefit from the laws of freedom of speech, despite freedom of speech being written in the constitution, and the average student studies longer hours at public and private schools. Furthermore, the average worker works longer hours per day compared to other OECD nations; they are also working in very hierarchically-structured organizations, such as the Korean conglomerate. Korean conglomerates are quasi mini-monarchies that function like miniature kingdoms within this

liberal democracy. Koreans seem to be unable to shed their feudal and totalitarian past. I would label South Korea as a capitalistic monarchy posing as a democracy. Indication of this is prevalent in society. One example is the strict hierarchy within these conglomerates. The conglomerate was once a family of privilege hand picked by the government that maneuvers its assigned company in its government-appointed industry, eg. LG appliances, Samsung electronics, Hyundai industries, etc. To this day, the conglomerates are structured in this way, as a monarches in themselves. These totalitarian-like entities, where the CEO is king, pressure employees in ways that are very autocratic.

The ROK is a very conservative nation. The minimum wage has only recently been adjusted to modern standards. Not long ago, it was roughly four USD per hour. The ROK has experienced economic hardships countless times due to both domestic and international influences. The so-called ‘miracle of the Han’, while experiencing rapid growth for the betterment of the nation, has also experienced numerous setbacks and negative side effects due to rapid advancement in its modern history.

The ROK has since become an ultra-capitalist democracy with a society that is overly conservative both socially and fiscally. While boasting the 13th largest economy in the world, this social lag can clearly be attributed to its expansive economic growth in such a short period of time. This is truly remarkable. This is a surreal first-world technology, third-world mentality.

Data Explanation

The first dataset, '101_DT_1B34E09_20240718123555.csv', contains data of leading causes of death by gender *(appendix 02.01) (appendix 02.02) (appendix 02.04)*, The second data set, '101 _DT _1B34E12 _20240718123805.csv', contains data of leading causes of death by geographic location *(appendix 02.01) (appendix 02.02) (appendix 02.05)*, and the third dataset, 'who_suicide_statistics.csv', contains data on international suicide rates conducted by the WHO *(appendix 02.01) (appendix 02.02)*.

Unfortunately the first two datasets do not contain suicide statistics. They include primarily data on deaths caused by disease *(appendix 02.02) (appendix 02.04)*. I will need to find additional datasets to compensate for this. The third dataset contains WHO statistics on suicide rates internationally; however, there are many countries that are missing from this dataset *(appendix 02.02) (appendix 02.05)*. I will need to find additional datasets with more complete data pertaining to suicide rates internationally *(appendix 02.10)*. I was, however, able to extract data on the Republic of Korea and other countries such as the United States, Russia, and Japan, which are all relevant for comparisons *(appendix 02.08)*.

Due to the gaps in data among the three datasets, I will include additional datasets that can compensate for the three inadequate datasets. I will conduct further research in the next milestone.

I added several datasets as dt01-dt03 table structures were inadequate *(appendix 03.01) (appendix 03.02) (appendix 03.03) (appendix 03.04) (appendix 03.05) (appendix 03.06)*.

Dataset five contains data on general life stress in the ROK, dataset six contains data on home stress in the ROK, dataset seven contains data on school stress in the ROK, dataset eight contains data on work stress in the ROK, dataset nine contains data on alcohol consumption of 19 year olds in the ROK, dataset 10 contains data on alcohol consumption of 20 year olds in the ROK, dataset 11 contains data on alcohol management of 19 year olds in the ROK, dataset 12 contains data on alcohol management of 20 year olds in the ROK, dataset 13 contains data on suicide impulse of people over the age of 13 in the ROK, dataset 14 contains data on suicide impulse of people over the age of 13 in the ROK, dataset 15 contains data on attempted suicides of people over the age of 65 in the ROK, dataset 16 contains data on attempted suicides of people over the age of 65 in the ROK, dataset 17 contains data on the rates of smoking and drinking of people over 19 years old in the ROK, dataset 18 contains data on the rates of smoking and drinking of people over 20 years old in the ROK, dataset 19 contains data on porn search categories, dataset 20 contains data on porn analysis categories, dataset 21 contains data on porn video categories, dataset 22 contains data on porn statistics worldwide, dataset 23 contains data on demographics of the ROK, dataset 24 contains data on demographics of the

ROK, dataset 25 contains data on economic sentiment in the ROK, dataset 26 contains data on feelings of sadness in the ROK, dataset 27 contains data on feelings of happiness in the ROK, dataset 28 contains data on happiness index in 2015 in the ROK, dataset 29 contains data on happiness index in 2016 in the ROK, dataset 30 contains data on happiness index in 2017 in the ROK, dataset 31 contains data on happiness index in 2018 in the ROK, dataset 32 contains data on happiness index in 2019 in the ROK, dataset 33 contains data on population by income level in the ROK, dataset 34 contains data on perceived stress levels in the ROK, dataset 35 contains data on reasons for committing suicide among people over 64 years old in the ROK, dataset 36 contains data on reasons for depression among people over 64 years old in the ROK, and dataset 37 contains data on the world happiness index (international) *(appendix 03.01) (appendix 03.02) (appendix 03.03) (appendix 03.04) (appendix 03.05) (appendix 03.06)*.

Methods

My target variables for various models are the suicide rate in Korea compared to other countries, specific age groups, geographic locations, and other demographic and economic factors.

I performed EDA and cleaned the datasets. Data cleaning was critical in organizing the two domestic datasets: leading causes of death by gender and leading causes of death by geographic location, for uniformity between

the two *(appendix 02.04) (appendix 02.05)*. I tested the three datasets for accuracy using linear and logistic regression models. I also used regression analysis to compare the two domesting datasets. I attempted to use time-series analysis on all three datasets to recognize trends over periods of time, but due to the structure of the datasets, I was unable to perform this analysis. I will likely need to implement other models for predictive analysis using classification and clustering models.

I performed EDA on the additional datasets for milestone three *(appendix 03.01) (appendix 03.02) (appendix 03.03) (appendix 03.04) (appendix 03.05) (appendix 03.06)*. I performed countless data transformations on almost all datasets. The transformations included changing data type, transposing, renaming columns for consistency, filling NaN values, replacing missing values, etc. Due to the structure of datasets dt05-dt18, I had to convert them from tuples to dictionaries and then transpose them. These datasets are from KOSIS, the Korean government's official sources, which tend to be unorganized and require a lot of cleaning. These datasets also required removal of strings in order for me to convert categorical data to floats for visualization. I also had to remove 'continuous' labels in some datasets.

```
# 10.02.05.01
# column rename to remove spaces
# dt04

dt04_combined_____________00_rn = dt04_combined_____________00.rename(columns = {
    'Region Name': 'name_region',
    'Country Name': 'name_country',
    'Year': 'year',
    'Sex': 'gender',
    '0 Year': '00',
    '1-4 Years': '01_04',
    '5-9 Years': '05_09',
    '10-14 Years': '10_14',
    '15-19 Years': '15_19',
    '20-24 Years': '20_24',
    '25-29 Years': '25_29',
    '30-34 Years': '30_34',
    '35-39 Years': '35_39',
    '40-44 Years': '40_44',
    '45-49 Years': '45_49',
    '50-54 Years': '50_54',
    '55-59 Years': '55_59',
    '60-64 Years': '60_64',
    '65-69 Years': '65_69',
    '70-74 Years': '70_74',
    '75-79 Years': '75_79',
    '80-84 Years': '80_84',
    '85+ Years': '85_up',
    'Unknown Age': 'age_unknown',
    'No of Suicides': 'num_suicide',
    'Percentage of cause-specific deaths out of total deaths': 'death_specific_total',
    'Death rate per 100 000 population': 'death_rate_100000'
    })
```

The above is a naming protocol that I use to organize the columns within my datasets. This naming convention makes working with the data much more convenient and efficient. I am able to select entire columns in an organized structure.

```
# 10.02.08.02
# change types string / float
# dt04

dt04_combined_____________00_rn_cv['name_region'].astype('string')
dt04_combined_____________00_rn_cv['name_country'].astype('string')
dt04_combined_____________00_rn_cv['year'].astype('int64')
dt04_combined_____________00_rn_cv['gender'].astype('string')
dt04_combined_____________00_rn_cv['00'].astype('float')
dt04_combined_____________00_rn_cv['01_04'].astype('float')
dt04_combined_____________00_rn_cv['05_09'].astype('float')
dt04_combined_____________00_rn_cv['10_14'].astype('float')
dt04_combined_____________00_rn_cv['15_19'].astype('float')
dt04_combined_____________00_rn_cv['20_24'].astype('float')
dt04_combined_____________00_rn_cv['25_29'].astype('float')
dt04_combined_____________00_rn_cv['30_34'].astype('float')
dt04_combined_____________00_rn_cv['35_39'].astype('float')
dt04_combined_____________00_rn_cv['40_44'].astype('float')
dt04_combined_____________00_rn_cv['45_49'].astype('float')
dt04_combined_____________00_rn_cv['50_54'].astype('float')
dt04_combined_____________00_rn_cv['55_59'].astype('float')
dt04_combined_____________00_rn_cv['60_64'].astype('float')
dt04_combined_____________00_rn_cv['65_69'].astype('float')
dt04_combined_____________00_rn_cv['70_74'].astype('float')
dt04_combined_____________00_rn_cv['75_79'].astype('float')
dt04_combined_____________00_rn_cv['80_84'].astype('float')
dt04_combined_____________00_rn_cv['85_up'].astype('float')
dt04_combined_____________00_rn_cv['age_unknown'].astype('float')
dt04_combined_____________00_rn_cv['num_suicide'].astype('float')
dt04_combined_____________00_rn_cv['death_specific_total'].astype('float')
dt04_combined_____________00_rn_cv['death_rate_100000'].astype('float')
```

In the above example, I changed types according to the models to be run on each dataset. I changed types of multiple datasets.

```
Cut this cell (X)
# 10.02.00.04
# replace symbols
# dt05

dt05_stress_general_____00_rn_dict_df_tp['2008_00'] = dt05_stress_general_____00_rn_dict_df_tp['2008_00'].replace(['-', '*', 'Total', 'Severe', 'Moderate', 'Weak'],
np.mean(pd.to_numeric(dt05_stress_general_____00_rn_dict_df_tp['2008_00'], errors='coerce')))

dt05_stress_general_____00_rn_dict_df_tp['2008_01'] = dt05_stress_general_____00_rn_dict_df_tp['2008_01'].replace(['-', '*', 'Total', 'Severe', 'Moderate', 'Weak'],
np.mean(pd.to_numeric(dt05_stress_general_____00_rn_dict_df_tp['2008_01'], errors='coerce')))

dt05_stress_general_____00_rn_dict_df_tp['2008_02'] = dt05_stress_general_____00_rn_dict_df_tp['2008_02'].replace(['-', '*', 'Total', 'Severe', 'Moderate', 'Weak'],
np.mean(pd.to_numeric(dt05_stress_general_____00_rn_dict_df_tp['2008_02'], errors='coerce')))

dt05_stress_general_____00_rn_dict_df_tp['2008_03'] = dt05_stress_general_____00_rn_dict_df_tp['2008_03'].replace(['-', '*', 'Total', 'Severe', 'Moderate', 'Weak'],
np.mean(pd.to_numeric(dt05_stress_general_____00_rn_dict_df_tp['2008_03'], errors='coerce')))

dt05_stress_general_____00_rn_dict_df_tp['2008_04'] = dt05_stress_general_____00_rn_dict_df_tp['2008_04'].replace(['-', '*', 'Total', 'Severe', 'Moderate', 'Weak'],
np.mean(pd.to_numeric(dt05_stress_general_____00_rn_dict_df_tp['2008_04'], errors='coerce')))

dt05_stress_general_____00_rn_dict_df_tp['2010_00'] = dt05_stress_general_____00_rn_dict_df_tp['2010_00'].replace(['-', '*', 'Total', 'Severe', 'Moderate', 'Weak'],
np.mean(pd.to_numeric(dt05_stress_general_____00_rn_dict_df_tp['2010_00'], errors='coerce')))

dt05_stress_general_____00_rn_dict_df_tp['2010_01'] = dt05_stress_general_____00_rn_dict_df_tp['2010_01'].replace(['-', '*', 'Total', 'Severe', 'Moderate', 'Weak'],
np.mean(pd.to_numeric(dt05_stress_general_____00_rn_dict_df_tp['2010_01'], errors='coerce')))

dt05_stress_general_____00_rn_dict_df_tp['2010_02'] = dt05_stress_general_____00_rn_dict_df_tp['2010_02'].replace(['-', '*', 'Total', 'Severe', 'Moderate', 'Weak'],
np.mean(pd.to_numeric(dt05_stress_general_____00_rn_dict_df_tp['2010_02'], errors='coerce')))

dt05_stress_general_____00_rn_dict_df_tp['2010_03'] = dt05_stress_general_____00_rn_dict_df_tp['2010_03'].replace(['-', '*', 'Total', 'Severe', 'Moderate', 'Weak'],
np.mean(pd.to_numeric(dt05_stress_general_____00_rn_dict_df_tp['2010_03'], errors='coerce')))

dt05_stress_general_____00_rn_dict_df_tp['2010_04'] = dt05_stress_general_____00_rn_dict_df_tp['2010_04'].replace(['-', '*', 'Total', 'Severe', 'Moderate', 'Weak'],
np.mean(pd.to_numeric(dt05_stress_general_____00_rn_dict_df_tp['2010_04'], errors='coerce')))

dt05_stress_general_____00_rn_dict_df_tp['2012_00'] = dt05_stress_general_____00_rn_dict_df_tp['2012_00'].replace(['-', '*', 'Total', 'Severe', 'Moderate', 'Weak'],
np.mean(pd.to_numeric(dt05_stress_general_____00_rn_dict_df_tp['2012_00'], errors='coerce')))

dt05_stress_general_____00_rn_dict_df_tp['2012_01'] = dt05_stress_general_____00_rn_dict_df_tp['2012_01'].replace(['-', '*', 'Total', 'Severe', 'Moderate', 'Weak'],
np.mean(pd.to_numeric(dt05_stress_general_____00_rn_dict_df_tp['2012_01'], errors='coerce')))

dt05_stress_general_____00_rn_dict_df_tp['2012_02'] = dt05_stress_general_____00_rn_dict_df_tp['2012_02'].replace(['-', '*', 'Total', 'Severe', 'Moderate', 'Weak'],
np.mean(pd.to_numeric(dt05_stress_general_____00_rn_dict_df_tp['2012_02'], errors='coerce')))

dt05_stress_general_____00_rn_dict_df_tp['2012_03'] = dt05_stress_general_____00_rn_dict_df_tp['2012_03'].replace(['-', '*', 'Total', 'Severe', 'Moderate', 'Weak'],
```

In the above example, I replaced symbols for certain datasets according to the model to be run on them. I did this with multiple datasets.

```
# 10.02.18.04
# replace symbols
# dt14

dt14_suicide_impulse_____00_rn_dict_df_tp['2008_00'] = dt14_suicide_impulse_____00_rn_dict_df_tp['2008_00'].replace(['-', '*', 'Total', 'Severe', 'Moderate', 'Weak', 'Drinking',
'- Less than once per month', '- 2~3 times per month', '- 1~2 times per week', '- 3~4 times per week', '- Almost everyday', 'No drinking', '- Used to but quit', '- Never', 'Eat
breakfast regularly-Do', 'Eat breakfast regularly-Dont', 'Get enough sleep(6~8 hours)-Do', 'Get enough sleep(6~8 hours)-Dont', 'Exercise regularly-Do', 'Exercise regularly-Dont', 'Get a
regular check up-Do', 'Get a regular check up-Dont', "Eat breakfast regularly-Don't", "Get enough sleep(6~8 hours)-Don't", "Exercise regularly-Don't", "Get a regular check up-Don't",
'Felt', 'subtotal', '-Economic hardships', ' -Problems with significant other', '-Physical/mental sickness depression or disability', '-Job Issues', '-Loneliness solitude', '-Family
issues', '-School performance concerns about higher education', '-Problems with friends or colleagues', '-Others', "Didn't feel"],
np.mean(pd.to_numeric(dt14_suicide_impulse_____00_rn_dict_df_tp['2008_00'], errors='coerce')))

dt14_suicide_impulse_____00_rn_dict_df_tp['2008_01'] = dt14_suicide_impulse_____00_rn_dict_df_tp['2008_01'].replace(['-', '*', 'Total', 'Severe', 'Moderate', 'Weak', 'Drinking',
'- Less than once per month', '- 2~3 times per month', '- 1~2 times per week', '- 3~4 times per week', '- Almost everyday', 'No drinking', '- Used to but quit', '- Never', 'Eat
breakfast regularly-Do', 'Eat breakfast regularly-Dont', 'Get enough sleep(6~8 hours)-Do', 'Get enough sleep(6~8 hours)-Dont', 'Exercise regularly-Do', 'Exercise regularly-Dont', 'Get a
regular check up-Do', 'Get a regular check up-Dont', "Eat breakfast regularly-Don't", "Get enough sleep(6~8 hours)-Don't", "Exercise regularly-Don't", "Get a regular check up-Don't",
'Felt', 'subtotal', '-Economic hardships', ' -Problems with significant other', '-Physical/mental sickness depression or disability', '-Job Issues', '-Loneliness solitude', '-Family
issues', '-School performance concerns about higher education', '-Problems with friends or colleagues', '-Others', "Didn't feel"],
np.mean(pd.to_numeric(dt14_suicide_impulse_____00_rn_dict_df_tp['2008_01'], errors='coerce')))

dt14_suicide_impulse_____00_rn_dict_df_tp['2008_02'] = dt14_suicide_impulse_____00_rn_dict_df_tp['2008_02'].replace(['-', '*', 'Total', 'Severe', 'Moderate', 'Weak', 'Drinking',
'- Less than once per month', '- 2~3 times per month', '- 1~2 times per week', '- 3~4 times per week', '- Almost everyday', 'No drinking', '- Used to but quit', '- Never', 'Eat
breakfast regularly-Do', 'Eat breakfast regularly-Dont', 'Get enough sleep(6~8 hours)-Do', 'Get enough sleep(6~8 hours)-Dont', 'Exercise regularly-Do', 'Exercise regularly-Dont', 'Get a
regular check up-Do', 'Get a regular check up-Dont', "Eat breakfast regularly-Don't", "Get enough sleep(6~8 hours)-Don't", "Exercise regularly-Don't", "Get a regular check up-Don't",
'Felt', 'subtotal', '-Economic hardships', ' -Problems with significant other', '-Physical/mental sickness depression or disability', '-Job Issues', '-Loneliness solitude', '-Family
issues', '-School performance concerns about higher education', '-Problems with friends or colleagues', '-Others', "Didn't feel"],
np.mean(pd.to_numeric(dt14_suicide_impulse_____00_rn_dict_df_tp['2008_02'], errors='coerce')))

dt14_suicide_impulse_____00_rn_dict_df_tp['2008_03'] = dt14_suicide_impulse_____00_rn_dict_df_tp['2008_03'].replace(['-', '*', 'Total', 'Severe', 'Moderate', 'Weak', 'Drinking',
'- Less than once per month', '- 2~3 times per month', '- 1~2 times per week', '- 3~4 times per week', '- Almost everyday', 'No drinking', '- Used to but quit', '- Never', 'Eat
breakfast regularly-Do', 'Eat breakfast regularly-Dont', 'Get enough sleep(6~8 hours)-Do', 'Get enough sleep(6~8 hours)-Dont', 'Exercise regularly-Do', 'Exercise regularly-Dont', 'Get a
regular check up-Do', 'Get a regular check up-Dont', "Eat breakfast regularly-Don't", "Get enough sleep(6~8 hours)-Don't", "Exercise regularly-Don't", "Get a regular check up-Don't",
'Felt', 'subtotal', '-Economic hardships', ' -Problems with significant other', '-Physical/mental sickness depression or disability', '-Job Issues', '-Loneliness solitude', '-Family
issues', '-School performance concerns about higher education', '-Problems with friends or colleagues', '-Others', "Didn't feel"],
np.mean(pd.to_numeric(dt14_suicide_impulse_____00_rn_dict_df_tp['2008_03'], errors='coerce')))
```

In the above example, I replaced symbols for certain datasets according to the model to be run on them. I did this with multiple datasets.

```
# 10.02.20.06
# replace NaN with mean
# dt16

dt16_suicide_reason_____00_rn_dict_df_tp.fillna(mean, inplace=True)
dt16_suicide_reason_____00_rn_dict_df_tp_ntf.fillna(mean, inplace=True)
```

```
# 10.02.21.01
# convert tuple to dict
# dt17

dt17_smoke_drink_19_____00_rn_dict = dict(dt17_smoke_drink_19_____00_rn)
```

```
# 10.02.21.02
# convert dict to df
# dt17

dt17_smoke_drink_19_____00_rn_dict_df = pd.DataFrame.from_dict(dt17_smoke_drink_19_____00_rn_dict, orient='index')
```

```
# 10.02.21.03
# transpose df
# dt17

dt17_smoke_drink_19_____00_rn_dict_df_tp = dt17_smoke_drink_19_____00_rn_dict_df.T
dt17_smoke_drink_19_____00_rn_dict_df_tp_ntf = dt17_smoke_drink_19_____00_rn_dict_df.T
```

In the above example, I replaced NaN values for certain datasets according to the model to be run on them. I did this with multiple datasets.

```
# 10.02.22.06
# replace NaN with mean
# dt18

dt18_smoke_drink_20_____00_rn_dict_df_tp.fillna(mean, inplace=True)
dt18_smoke_drink_20_____00_rn_dict_df_tp_ntf.fillna(mean, inplace=True)
```

```
# 10.02.23.01
# replace NaN with mean
# dt19

dt19_ph_categories_____00_rn.fillna(mean, inplace=True)
```

```
# 10.02.24.01
# replace NaN with mean
# dt20

'''dt20_ph_analysis_____00_rn.fillna(mean, inplace=True)'''
```

```
'dt20_ph_analysis_____00_rn.fillna(mean, inplace=True)'
```

```
# 10.02.25.01
# convert to date-time
# dt21

ts, ms = '20.12.2016 09:38:42,76'.split(',')
dt21_ph_videos_____00_rn['date_pub'] = datetime.strptime(ts, '%d.%m.%Y %H:%M:%S')
```

```
# 10.02.25.02
# replace NaN with mean
# dt21

dt21_ph_videos_____00_rn.fillna(mean, inplace=True)
```

```
# 10.02.26.01
# replace NaN with mean
# dt21

'''dt22_ph_tot_____00_rn.fillna(mean, inplace=True)'''
```

In the above example, I replaced NaN values for certain datasets according to the model to be run on them. I did this with multiple datasets.

```
# 10.02.28.01
# preparing data for modeling
# add index column
# index column to select specific rows
# dt04

dt04_combined_____________00_rn_cv.insert(0, 'index', range(0, 0 + len(dt04_combined_____________00_rn_cv)))
```

```
# 10.02.28.02
# preparing data for modeling
# create dummy variables
# due to returning boolean values, converting dummmies to integers
# dt04

dt04_combined_____________00_rn_cv_lnr_dv01 = pd.get_dummies(dt04_combined_____________00_rn_cv, drop_first = True, dtype = int)
```

```
# 10.02.28.03
# preparing data for modeling
# split data
# select columns
# target variable: at birth
# dt04

dt04_lnr_x01 = dt04_combined_____________00_rn_cv_lnr_dv01.drop(['00'], axis = 1)
dt04_lnr_y01 = dt04_combined_____________00_rn_cv_lnr_dv01['00']
```

```
# 10.02.28.04
# preparing data for modeling
# split into train and test
# target variable: at birth
# dt04

dt04_lnr_x01_trn, dt04_lnr_x01_tst, dt04_lnr_y01_trn, dt04_lnr_y01_tst = train_test_split(dt04_lnr_x01, dt04_lnr_y01, test_size = 0.3, random_state = 0)
```

```
# 10.02.28.05
# preparing data for modeling
# assign regression variable
# target variable: at birth
# dt04

dt04_lnr_lr01 = LinearRegression()
```

In the above example, I performed regression models by replacing the target variable and creating dummies.

```
# 10.02.36.09
# run model
# return rmse and r2 dt04
# rmse: 3.19
# r2: 1.0
# target variable: number of suicides
# dt04

print(f'rmse: {dt04_lnr_rmse21}')
print(f'r2: {dt04_lnr_r221}')
```

```
rmse: 3.5276599611628754e-12
r2: 1.0
```

```
# 10.02.36.10
# replace NA with mode
# this step improves the accuracy of linear and logistic regressions
# dt04

dt04_combined_____________00_rn_cv['00']=dt04_combined_____________00_rn_cv['00'].fillna(dt04_combined_____________00_rn_cv['00'].mode()[0])
dt04_combined_____________00_rn_cv['01_04']=dt04_combined_____________00_rn_cv['01_04'].fillna(dt04_combined_____________00_rn_cv['01_04'].mode()[0])
dt04_combined_____________00_rn_cv['05_09']=dt04_combined_____________00_rn_cv['05_09'].fillna(dt04_combined_____________00_rn_cv['05_09'].mode()[0])
dt04_combined_____________00_rn_cv['10_14']=dt04_combined_____________00_rn_cv['10_14'].fillna(dt04_combined_____________00_rn_cv['10_14'].mode()[0])
dt04_combined_____________00_rn_cv['15_19']=dt04_combined_____________00_rn_cv['15_19'].fillna(dt04_combined_____________00_rn_cv['15_19'].mode()[0])
dt04_combined_____________00_rn_cv['20_24']=dt04_combined_____________00_rn_cv['20_24'].fillna(dt04_combined_____________00_rn_cv['20_24'].mode()[0])
dt04_combined_____________00_rn_cv['25_29']=dt04_combined_____________00_rn_cv['25_29'].fillna(dt04_combined_____________00_rn_cv['25_29'].mode()[0])
dt04_combined_____________00_rn_cv['30_34']=dt04_combined_____________00_rn_cv['30_34'].fillna(dt04_combined_____________00_rn_cv['30_34'].mode()[0])
dt04_combined_____________00_rn_cv['35_39']=dt04_combined_____________00_rn_cv['35_39'].fillna(dt04_combined_____________00_rn_cv['35_39'].mode()[0])
dt04_combined_____________00_rn_cv['40_44']=dt04_combined_____________00_rn_cv['40_44'].fillna(dt04_combined_____________00_rn_cv['40_44'].mode()[0])
dt04_combined_____________00_rn_cv['45_49']=dt04_combined_____________00_rn_cv['45_49'].fillna(dt04_combined_____________00_rn_cv['45_49'].mode()[0])
dt04_combined_____________00_rn_cv['50_54']=dt04_combined_____________00_rn_cv['50_54'].fillna(dt04_combined_____________00_rn_cv['50_54'].mode()[0])
dt04_combined_____________00_rn_cv['55_59']=dt04_combined_____________00_rn_cv['55_59'].fillna(dt04_combined_____________00_rn_cv['55_59'].mode()[0])
dt04_combined_____________00_rn_cv['60_64']=dt04_combined_____________00_rn_cv['60_64'].fillna(dt04_combined_____________00_rn_cv['60_64'].mode()[0])
dt04_combined_____________00_rn_cv['65_69']=dt04_combined_____________00_rn_cv['65_69'].fillna(dt04_combined_____________00_rn_cv['65_69'].mode()[0])
dt04_combined_____________00_rn_cv['70_74']=dt04_combined_____________00_rn_cv['70_74'].fillna(dt04_combined_____________00_rn_cv['70_74'].mode()[0])
dt04_combined_____________00_rn_cv['75_79']=dt04_combined_____________00_rn_cv['75_79'].fillna(dt04_combined_____________00_rn_cv['75_79'].mode()[0])
dt04_combined_____________00_rn_cv['80_84']=dt04_combined_____________00_rn_cv['80_84'].fillna(dt04_combined_____________00_rn_cv['80_84'].mode()[0])
dt04_combined_____________00_rn_cv['85_up']=dt04_combined_____________00_rn_cv['85_up'].fillna(dt04_combined_____________00_rn_cv['85_up'].mode()[0])
dt04_combined_____________00_rn_cv['age_unknown']=dt04_combined_____________00_rn_cv['age_unknown'].fillna(dt04_combined_____________00_rn_cv['age_unknown'].mode()[0])
dt04_combined_____________00_rn_cv['num_suicide']=dt04_combined_____________00_rn_cv['num_suicide'].fillna(dt04_combined_____________00_rn_cv['num_suicide'].mode()[0])
dt04_combined_____________00_rn_cv['death_specific_total']=dt04_combined_____________00_rn_cv['death_specific_total'].fillna(dt04_combined_____________00_rn_cv['death_specific_total'].mode()[0])
dt04_combined_____________00_rn_cv['death_rate_100000']=dt04_combined_____________00_rn_cv['death_rate_100000'].fillna(dt04_combined_____________00_rn_cv['death_rate_100000'].mode()[0])
```

The above example are ways in which I transformed the datasets according to the models to be run on them.

```
# 10.02.40.01
# preparing data for modeling
# create dummy variables
# due to returning boolean values, converting dummmies to integers
# dt04

dt04_combined_____________00_rn_lgr_dv00 = pd.get_dummies(dt04_combined_____________00_rn, drop_first = True, dtype = int)
```

```
# 10.02.40.02
# preparing data for modeling
# split data
# select columns
# target variable: age 00
# dt04

dt04_lgr_x00 = dt04_combined_____________00_rn_lgr_dv00.drop('00', axis=1)
dt04_lgr_y00 = dt04_combined_____________00_rn_lgr_dv00['00']
```

```
# 10.02.40.03
# preparing data for modeling
# assign regression variable
# target variable: age 00
# dt04

dt04_lgr_lr00 = LogisticRegression()
```

```
# 10.02.40.04
# preparing data for modeling
# split into train and test
# target variable: age 00
# dt04

dt04_lgr_x00_trn, dt04_lgr_x00_tst, dt04_lgr_y00_trn, dt04_lgr_y00_tst = train_test_split(dt04_lgr_x00, dt04_lgr_y00, test_size = 0.2, random_state = 42)
```

In the above example, I performed regression models by replacing the target variable and creating dummies.

```
# 10.02.59.01
# replace NaN with mean
# dt23

dt23_408_03_____________00_rn.fillna(mean, inplace=True)
```

```
# 10.02.59.02
# replace NaN with mean
# dt24

dt24_408_04_____________00_rn.fillna(mean, inplace=True)
```

```
# 10.02.59.03
# replace NaN with mean
# dt25

dt25_index_eco_sent_____00_rn.fillna(mean, inplace=True)
```

```
# 10.02.59.04
# replace NaN with mean
# dt26

dt26_sadness_____________00_rn.fillna(mean, inplace=True)
```

```
# 10.02.59.05
# replace NaN with mean
# dt27

dt27_happiness_01________00_rn.fillna(mean, inplace=True)
```

```
# 10.02.59.06
# replace NaN with mean
# dt28

dt28_happiness_2015______00_rn.fillna(mean, inplace=True)
```

```
# 10.02.59.07
# replace NaN with mean
# dt29

dt29_happiness_2016______00_rn.fillna(mean, inplace=True)
```

The above example are ways in which I transformed the datasets according to the models to be run on them.

```
# 10.02.71.07
# fit data for modeling
# fit variables to model
# target variable: index_happiness_2022
# dt27

dt27_lgr_y01_pdct = dt27_lgr_lr01.predict(dt27_lgr_x01_tst)
```

```
# 10.02.71.08
# assign variable for accuracy
# return accuracy
# accuracy: 0.926
# target variable: index_happiness_2022
# dt27

dt27_lgr_acc01 = accuracy_score(dt27_lgr_y01_tst, dt27_lgr_y01_pdct)
print('Accuracy:', dt27_lgr_acc01)
```

```
Accuracy: 0.9259259259259259
```

```
# 10.02.71.09
# return classification report
# target variable: index_happiness_2022
# dt27

print('Classification Report:')
print(classification_report(dt27_lgr_y01_tst, dt27_lgr_y01_pdct))
```

```
Classification Report:
              precision    recall  f1-score   support

           3       0.00      0.00      0.00         1
           4       0.75      1.00      0.86         6
           5       1.00      0.83      0.91         6
           6       1.00      1.00      1.00        12
           7       1.00      1.00      1.00         2

    accuracy                           0.93        27
   macro avg       0.75      0.77      0.75        27
weighted avg       0.91      0.93      0.91        27
```

In the above example, I performed regression models by replacing the target variable and creating dummies.

I used linear and logistic regression models to test for accuracy after data transformation. Some datasets were too inadequate to return any meaningful statistics. I also attempted to use Spark to extract data from a larger dataset of 20 gigabytes; unfortunately, I was unable to extract. I also attempted to conduct pickle serialization but was unable to process due to version discrepancies *(appendix 03.03)*.

Analysis

The box plot of dt01 shows how incomplete this dataset is *(appendix 02.03)*. This is due to the constraint in the timeline of the data. I will need to find similar data with a longer time period. In the second visualization for dt01, is a Gantt chart showing the number of deaths per cause *(appendix 02.04) (appendix 02.05)*. This data cover primarily diseases with an ambiguous 'external causes' column. The second dataset, dt02, is comparable to dt01 *(appendix 02.06) (appendix 02.07)*. For the third dataset, dt03, is a visualization of a bubble chart showing number of suicides internationally *(appendix 02.08)*. The United States and Russia are among the largest which is expected as they have among the largest populations. Also provided is a treemap of the same data for better visualization *(appendix 02.09)*. *(appendix 02.10)* is an area chart of the same data. Taking into consideration 'per capita' statistics, which are unavailable in these datasets, I provided some visualizations regarding populations in a bubble chart, area chart, and stacked bar chart *(appendix 02.11) (appendix 02.12) (appendix 02.13)*. The US and Russia have the highest populations which is reflective of their suicide rates. This data does not include China and India, which both have higher populations. I will need to find additional data to account for this. *(appendix 02.13)* also contains data on suicide rates by gender per country. In every instance, males, by far, outnumber females. The gender difference is also visualized in a pie chart with roughly three

quarters of the population being male *(appendix 02.14)*. For dt03, the next visualization is a bubble chart of suicides by age internationally *(appendix 02.15)*. According to this visualization, the age group 35-54 years has the highest rate of suicides with 55-74 years closely following in second. A bar chart of the same data also reflects these same results *(appendix 02.16)*. For dt03, I visualized suicides by year with a line graph. The graph covers the period of 1978 to 2016 *(appendix 02.17)*. Lastly is a line graph showing suicides from the 1980s to the present day for every country included in this dataset. It shows a major incline in the ROK compared to other countries *(appendix 02.18)*. I will need to find additional data to accommodate per capita variables for every country as the third dataset, dt03, has missing gaps of relevant data.

The box plot of dt25 shows almost zero outliers after transforming this dataset *(appendix 03.07)*. The box plot of dt27 shows zero outliers after transforming this dataset *(appendix 03.08)*. The box plot of dt28 shows few outliers after transforming this dataset *(appendix 03.09)*. The box plot of dt29 shows few outliers after transforming this dataset *(appendix 03.10)*. The box plot of dt30 shows few outliers after transforming this dataset *(appendix 03.11)*. The box plot of dt31 shows few outliers after transforming this dataset *(appendix 03.12)*. The box plot of dt32 shows few outliers after transforming this dataset *(appendix 03.13)*. There are visualizations of suicides by country; however, this is not measure per capita but rather total

suicides. This is reflected in the area maps with countries with higher populations having the most suicides *(appendix 03.15) (appendix 03.16) (appendix 03.17) (appendix 03.18)*. Suicides by region are illustrated with Europe having the highest numbers *(appendix 03.19) (appendix 03.20) (appendix 03.21)*. Males far outnumber females in suicides by more than double *(appendix 03.23) (appendix 03.24)*. The suicide rate of the ROK has steadily increased since the early 1980s *(appendix 03.25)*. Age seems to be the number-one factor in suicides in the ROK as there is a much higher rate among people older than 64 years. A logical explanation for this would be economic factors.

The ROK ranks quite low among OECD countries in the happiness index *(appendix 03.70) (appendix 03.71)*. There aren't any economic indicators to suggest that the economic situation of the ROK correlates with the high suicide rate. It could be assumed, however, that lower income households are those with inhabitants that are generally older than 64.

Conclusion

I can conclude that both Japan and the ROK have much higher rates of suicide compared to other countries. I can also conclude that it is likely that the ROK and Japan are among the highest in suicide rates per capita. According to the data, the US and Russia have the highest rates of suicide in total. However, it should be mentioned that Japan and Korea both have

significantly lower populations with Japan at roughly 130 million and the ROK with around 60 million compared to the populations of the US with roughly 340 million and Russia with around 150 million. I can also conclude that the number of suicides committed by males is significantly higher than that of their female counterparts; this is true for every country included in this dataset. Based on domestic data of the ROK, the age group that rates the highest in suicides correlates with the international values of the same data. According to the data, the age group with the highest suicide rate is people older than 64. This likely correlates with economic factors. The data was not adequate and relevant datasets weren't obtainable regarding the economic factors for this age group.

I would also assume that LGBT intolerance in the ROK also is a major contributor to the high suicide rate. I attempted to access relevant data to shed light on this topic, but I was unable to obtain. This will require further research. The stress levels of most age groups in the ROK were above average compared to other OECD countries.

Assumptions

It can be assumed that based on per-capita statistics, the ROK has the highest suicide rate with Japan coming in second. Based on the age groups with higher suicide rates, it is likely that the culprits are economically and demographically related. It is likely that there are universal societal reasons

for the gender imbalance of the worldwide suicide rate. It is also safe to assume that the fluctuations in the suicide rate over time can be attributed to economic trends over the same time periods.

Limitations

All three datasets used in milestone two are inadequate. The first two datasets, dt01 and dt02, primarily contain data relating to deaths caused by disease with one ambiguous column of 'external factors'. The third dataset, dt03, contains data regarding worldwide statistics on the issue of suicide. This dataset contains all necessary metrics outside of 'per-capita rates'. This dataset also does not include several countries that are directly relevant: China and India. There are several other nations that should be included in this research that are not included in this dataset. While the third dataset contains statistics on the rates of suicide per region, age, and gender, it does not contain any data regarding the causes. It also does not include other metrics that should be addressed such as demographically, economically, and socially related metrics.

Challenges

In milestone one, the previously-mentioned ethical considerations could pose problems in my data analysis. When it comes to the topic of suicide, there are so many potential metrics to consider. Lack of

consideration of more relevant metrics could lead to inaccuracies in my analysis. This could have an adverse effect on my research into prevention. Metrics such as marital status, age, economic situation, religion, personal relationship (family), gender, political affiliation, mental and physical health, geographic location, substance abuse, level of education, sexual orientation, and various other environmental and demographic aspects could be considered. In the case of the ROK, homosexuality is still considered a social taboo. I do believe that Korea's intolerance of LGBT significantly contributes to the suicide rate. Due to lack of data on the topic domestically, I will need to look to outside sources. This could raise some challenges as it is likely this type of data is not abundantly available. The Korean government censors internet content especially on Korean websites and search engines. Censorship is another obstacle that I will need to overcome.

In the second milestone, I did not detect any controversial statistics that could have ethical implications. However the data was limited and did not go in depth on the topic. In the obtaining of additional data for the next milestone, I should be aware of ways the above-mentioned metrics could be skewed and inaccurate.

Future Uses / Additional Applications

Once I have built a model, I could apply this to any region, particularly regions with abnormally high rates of suicide. I hope to build robust models

that can accommodate a variety of data types. I plan to incrementally make improvements to my models to achieve this. One fine tuned, I will further my research into regions of interest, such as the US, Russia, and Japan, which all have high suicide rates. My models will be tuned to statistically extract the causes of the issue and provide solutions with the highest probability of accuracy and success.

I have fine-tuned models that I used throughout this research project. I can confidently say that I could apply them, with very few adjustments, to fit most data types. I plan to use these models in future research and in my work.

Recommendations / Implementation

I will consider using other methods to test the accuracy of the data, assuming I find the specific data required for this research. I hope to apply predictive modeling techniques such as linear and logistic regression and categorical grouping of data by using ARIMA, K-Means clustering, and PCA. I also plan to obtain time-series data for predictive modeling. I want to implement ways of forecasting through models such as linear regression, moving averages, and time series.

Ethical Assessment

In the second milestone, I was unable to gather qualitative data. For the next milestone, I plan to gather more qualitative data to compensate and fill the missing gaps of my initial datasets. I should be aware of the ethical implications when implementing them. Using methods such as surveys for research into prevention, the parameters for participants' categorization into 'potential committer' must be taken into consideration. These parameters must be the same across the board. This raises the issue of what societal and economic metrics should be included in setting these parameters. For example, what is considered a social norm in one society may not be the case in another. Also, economic status among nations could raise ethical questions. To what degree does or should personal economic situations have an effect among societies on the topic. Happiness index should also be considered a potential issue in ethics. For example, what metrics are used in calculating the happiness index, and is this index used consistently across the board?

In the case of the ROK, which is highly intolerant of LGBT rights, there are immense ethical issues to be considered in this research. As mentioned in my assumptions, it is likely that the ROK has a much higher LGBT population than the statistics that are publicized. I will need to obtain data with metrics that can help me support this theory. The Korean government censors information of this nature.

References

2022, 2024.07.24, Deaths and death rates by cause(104 item)/By sex/By age(five-year age)
https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1B34E01&conn_path=I2&language=en

2022.12, 2024.07.24, Deaths by cause(104 item)/By sex
https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1B34E02&conn_path=I2&language=en

Basic historical (1979-2016) data by country, year and demographic groups
https://www.kaggle.com/datasets/szamil/who-suicide-statistics

CCO 1.0 Universal
https://www.kaggle.com/datasets/usamabuttar/world-happiness-report-2005-present
https://creativecommons.org/publicdomain/zero/1.0/

The Global Economy
https://www.theglobaleconomy.com/rankings/happiness/

2017, 2024.08.01, General Health Screening Beneficiaries and Statistics on the Number of Examinees by District
https://kosis.kr/statHtml/statHtml.do?orgId=350&tblId=DT_35007_N001&conn_path=I2&language=en

2022, 2024.08.01, Population by income level
https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1LC0021&conn_path=I2&language=en

2023, 2024.08.01, Feeling sad or hopeless
https://kosis.kr/statHtml/statHtml.do?orgId=177&tblId=DT_117_12_Y068&conn_path=I2&language=en

2023, 2024.08.01, Perceived stress
https://kosis.kr/statHtml/statHtml.do?orgId=177&tblId=DT_117_12_Y067&conn_path=I2&language=en

2020, 2024.08.01, Reason and Attempt to Think Suicide by General Feature of older persons(Over 65 Years Old)
https://kosis.kr/statHtml/statHtml.do?orgId=117&tblId=DT_117071_013&conn_path=I2&language=en

2024.06, 2024.08.01, Housing Sales Price Index by Scale (2021.6=100.0) https://kosis.kr/statHtml/statHtml.do?orgId=408&tblId=DT_40803_N0003&conn_path=I2&language=en

2024.06, 2024.08.01, Housing Jeonse Price Index by Scale (2021.6=100.0) https://kosis.kr/statHtml/statHtml.do?orgId=408&tblId=DT_40803_N0004&conn_path=I2&language=en

2024.07, 2024.08.01, Economic Sentiment Index https://kosis.kr/statHtml/statHtml.do?orgId=301&tblId=DT_513Y001&conn_path=I2&language=en

2020, 2024.08.01, Symptom of Depression by General Feature of older persons(Over 65 Years Old)

Appendix

02.01

```
# 08.02.01.01
# read csv
# assign variable
# dt01

dt01_death_cause_gend________00 = pd.read_csv('101_DT_1B34E09_20240718123555.csv')
```

```
# 08.02.01.02
# read csv
# assign variable
# dt02

dt02_death_cause_geo_________00 = pd.read_csv('101_DT_1B34E12_20240718123805.csv')
```

```
# 08.02.01.03
# read csv
# assign variable
# dt03

dt03_who_suicide_____________00 = pd.read_csv('who_suicide_statistics.csv')
```

02.02

```
# 09.02.02.01
# confirm column names
# dt01

dt01_death_cause_gend________00.columns
```

```
Index(['By the cause of death(104 items)', 'By gender', 'By province', 'Item',
       'UNIT', '1983 Year', '1984 Year', '1985 Year', '1986 Year', '1987 Year',
       '1988 Year', '1989 Year', '1990 Year', '1991 Year', '1992 Year',
       '1993 Year', '1994 Year', '1995 Year', '1996 Year', '1997 Year',
       '1998 Year', '1999 Year', '2000 Year', '2001 Year', '2002 Year',
       '2003 Year', '2004 Year', '2005 Year', '2006 Year', '2007 Year',
       '2008 Year', '2009 Year', '2010 Year', '2011 Year', '2012 Year',
       '2013 Year', '2014 Year', '2015 Year', '2016 Year', '2017 Year',
       '2018 Year', '2019 Year', '2020 Year', '2021 Year', '2022 Year',
       'Unnamed: 45'],
      dtype='object')
```

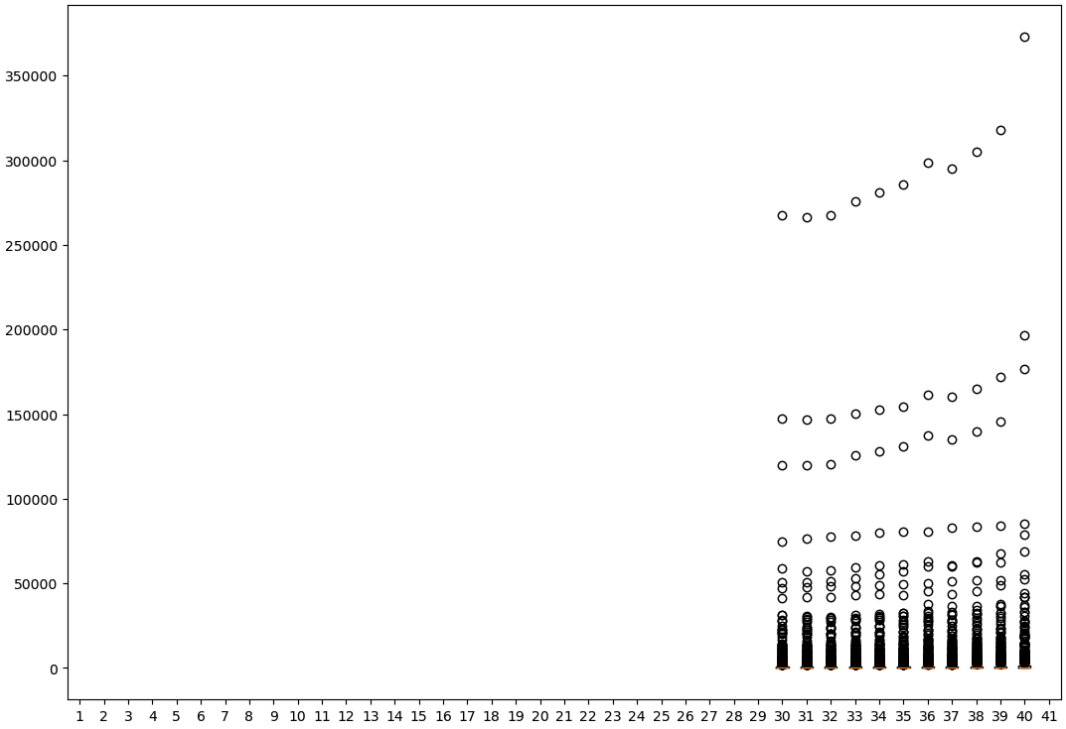
```
# 09.02.02.02
# confirm column names
# dt02

dt02_death_cause_geo_________00.columns
```

```
Index(['By the cause of death(104 items)', 'By province', 'By gender', 'Item',
       'UNIT', '1996 Year', '1997 Year', '1998 Year', '1999 Year', '2000 Year',
       '2001 Year', '2002 Year', '2003 Year', '2004 Year', '2005 Year',
       '2006 Year', '2007 Year', '2008 Year', '2009 Year', '2010 Year',
       '2011 Year', '2012 Year', '2013 Year', '2014 Year', '2015 Year',
       '2016 Year', '2017 Year', '2018 Year', '2019 Year', '2020 Year',
       '2021 Year', '2022 Year', 'Unnamed: 32'],
      dtype='object')
```

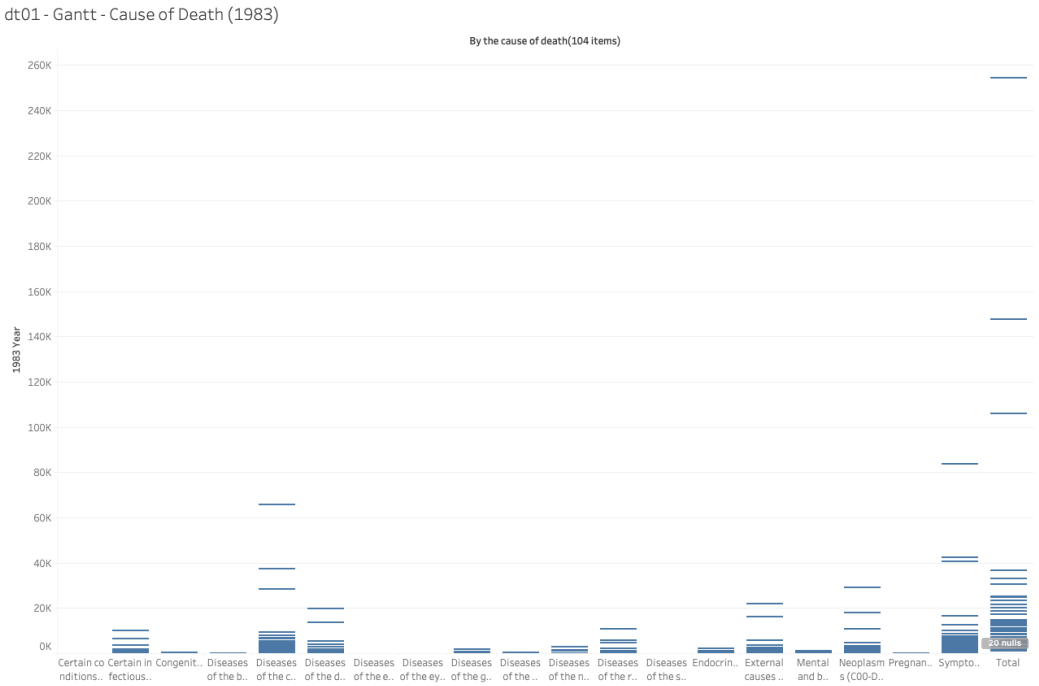
```
# 09.02.02.03
# confirm column names
# dt03

dt03_who_suicide_____________00.columns
```

```
Index(['country', 'year', 'sex', 'age', 'suicides_no', 'population'], dtype='object')
```

02.03

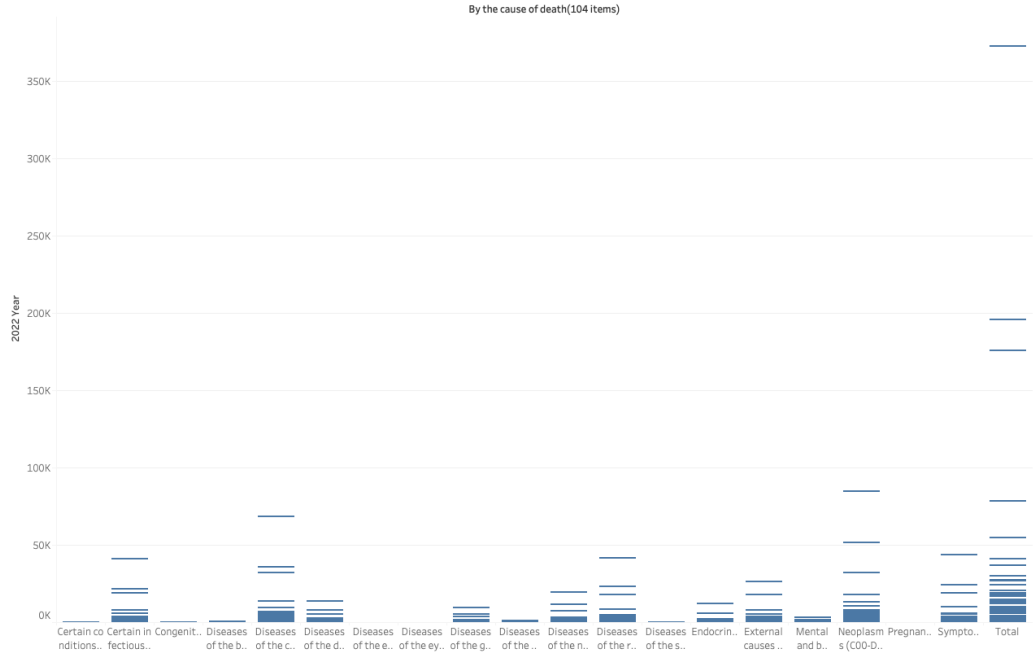

350000
300000
250000
200000
150000
100000
50000
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41

02.04


dt01 - Gantt - Cause of Death (1983)
By the cause of death(104 items)
1983 Year
260K
240K
220K
200K
180K
160K
140K
120K
100K
80K
60K
40K
20K
0K
20 nulls
Certain conditions..
Certain infectious..
Congenit..
Diseases of the b..
Diseases of the c..
Diseases of the d..
Diseases of the e..
Diseases of the ey..
Diseases of the g..
Diseases of the ..
Diseases of the n..
Diseases of the r..
Diseases of the s..
Endocrin..
External causes ..
Mental and b..
Neoplasms (C00-D..
Pregnan..
Sympto..
Total

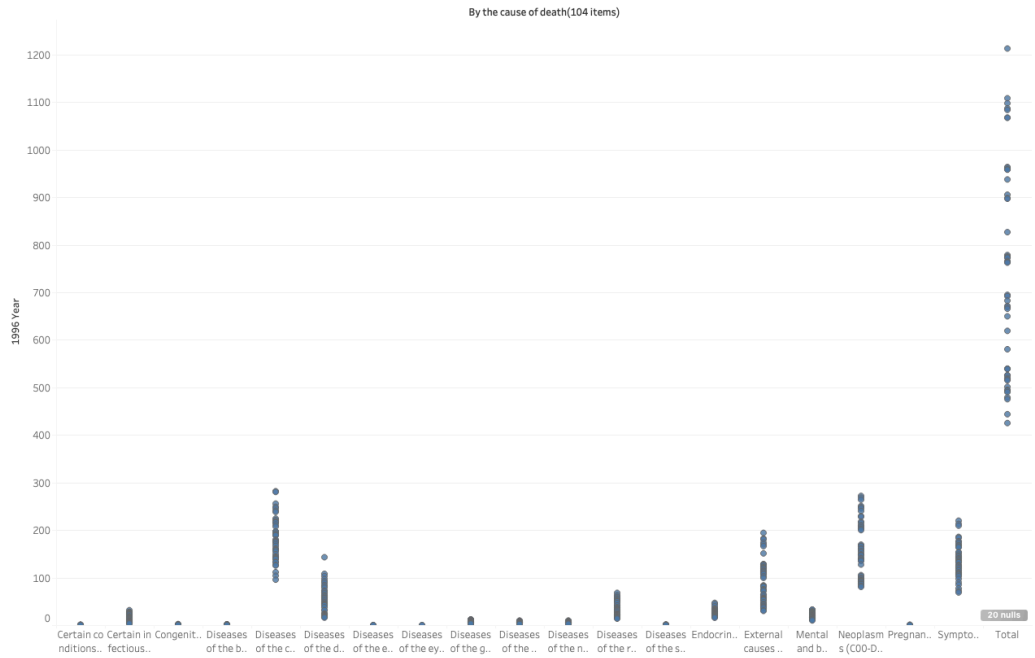
02.05

dt01 - Gantt - Cause of Death (2022)

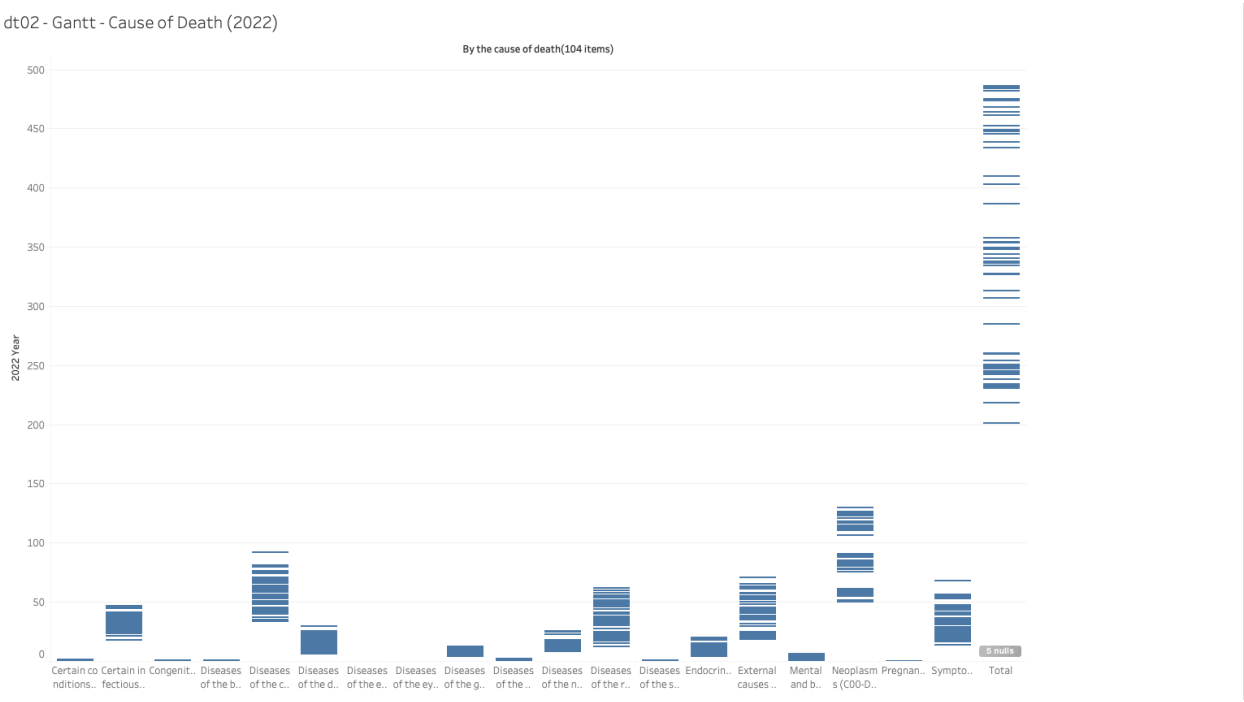


02.06

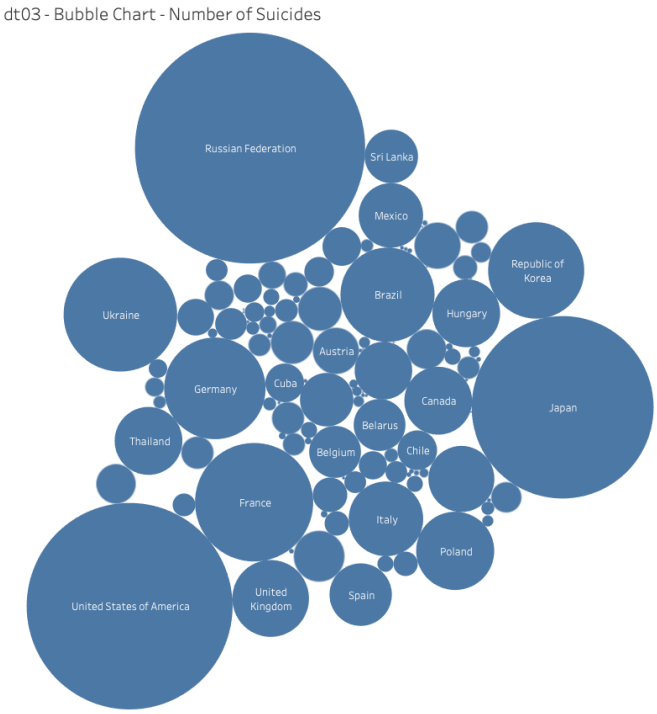
dt02 - Gantt - Cause of Death (1996)



02.07

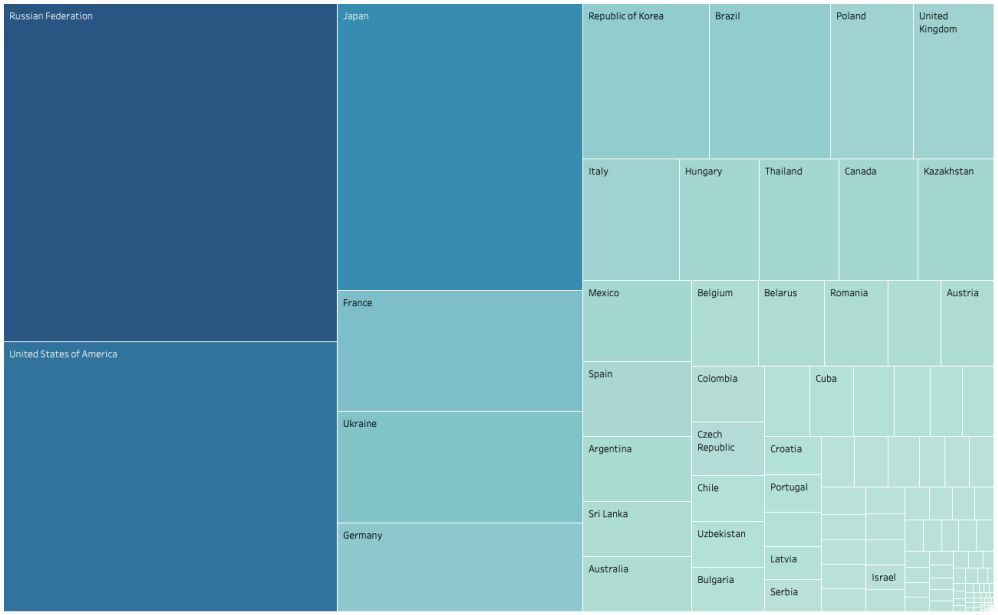

dt02 - Gantt - Cause of Death (2022)
By the cause of death(104 items)
2022 Year
500
450
400
350
300
250
200
150
100
50
0
5 nulls
Certain co nditions..
Certain in fectious..
Congenit..
Diseases of the b..
Diseases of the c..
Diseases of the d..
Diseases of the e..
Diseases of the ey..
Diseases of the g..
Diseases of the ..
Diseases of the n..
Diseases of the r..
Diseases of the s..
Endocrin..
External causes ..
Mental and b..
Neoplasm s (C00-D..
Pregnan..
Sympto..
Total

02.08


dt03 - Bubble Chart - Number of Suicides
Russian Federation
Sri Lanka
Mexico
Republic of Korea
Brazil
Hungary
Ukraine
Austria
Germany
Cuba
Canada
Japan
Belarus
Thailand
Belgium
Chile
France
Italy
Poland
United States of America
United Kingdom
Spain

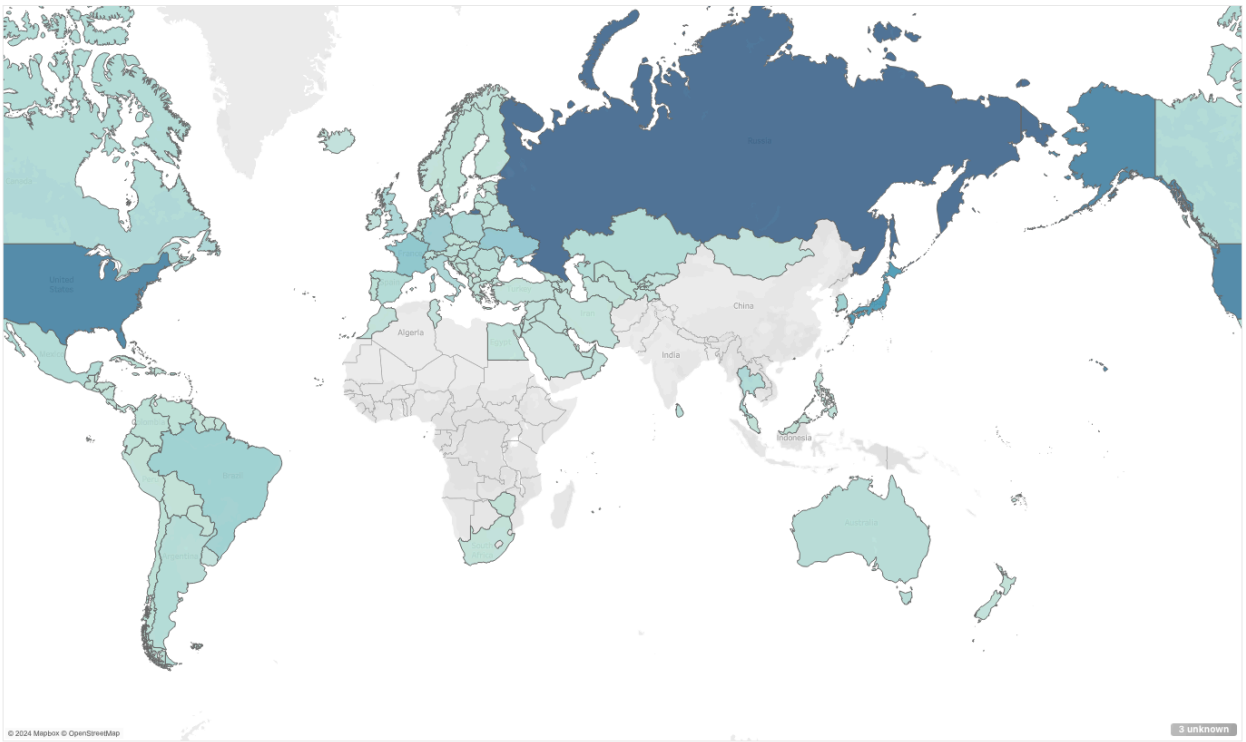
02.09

dt03 - Tree Map - Number of Suicides

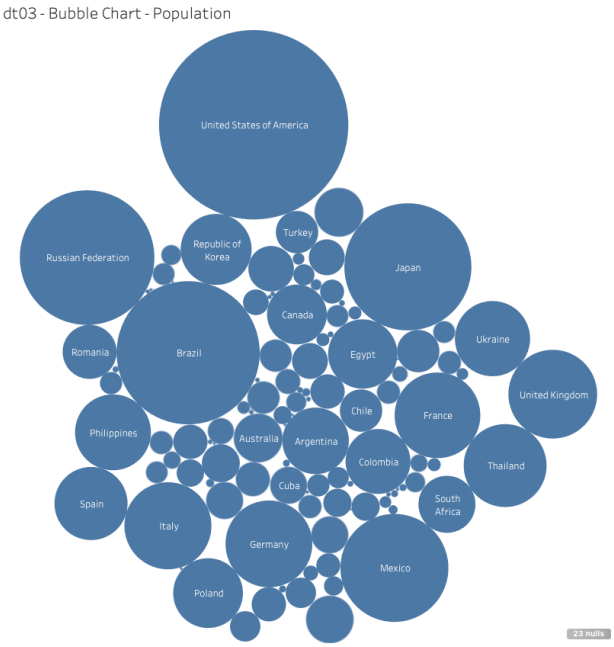


02.10

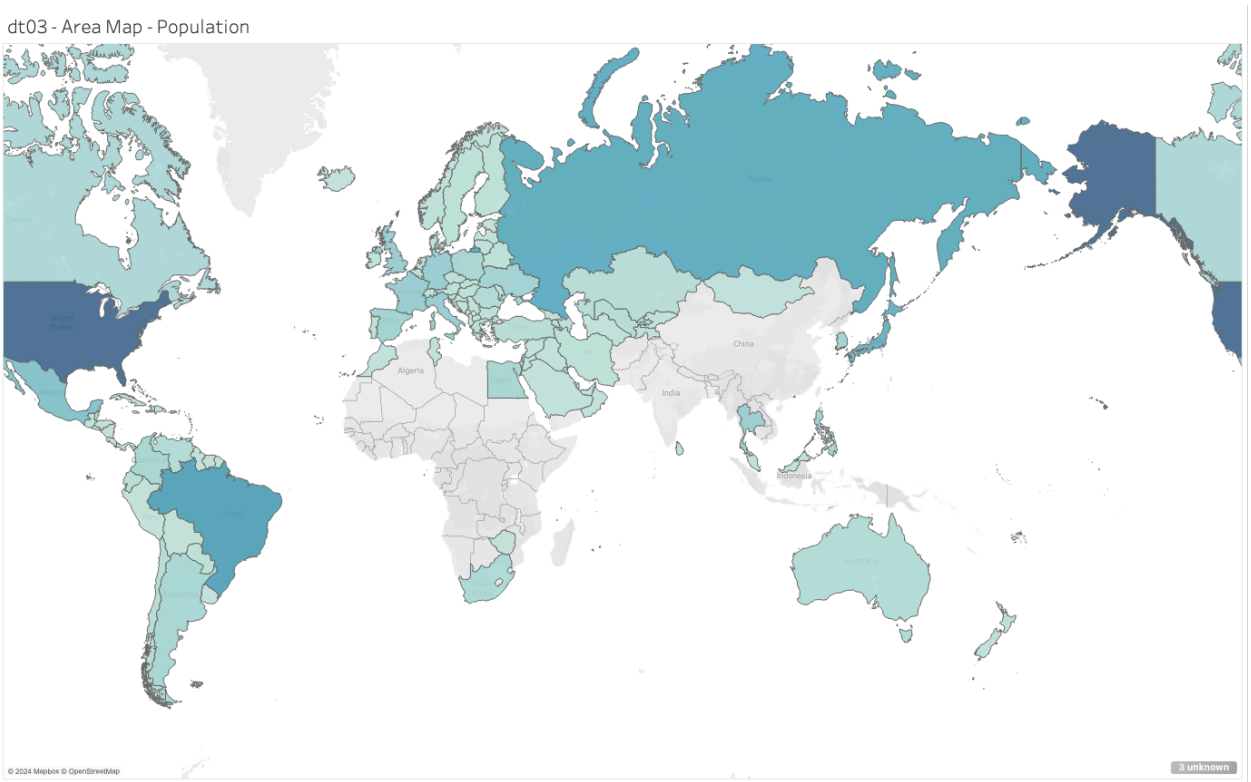
dt03 - Area Map - Number of Suicides



02.11

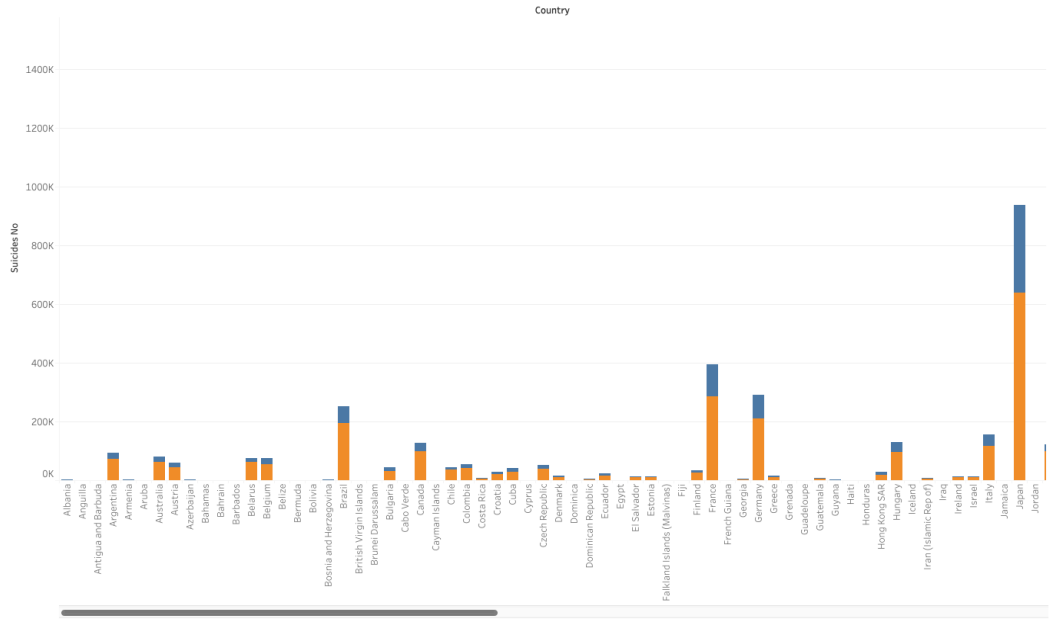

dt03 - Bubble Chart - Population
United States of America
Russian Federation
Republic of Korea
Turkey
Japan
Canada
Brazil
Romania
Egypt
Ukraine
United Kingdom
Chile
France
Philippines
Australia
Argentina
Colombia
Thailand
Cuba
Spain
South Africa
Italy
Germany
Mexico
Poland
23 nulls

02.12


dt03 - Area Map - Population
Russia
China
India
Algeria
Indonesia
Australia
© 2024 Mapbox © OpenStreetMap
3 unknown

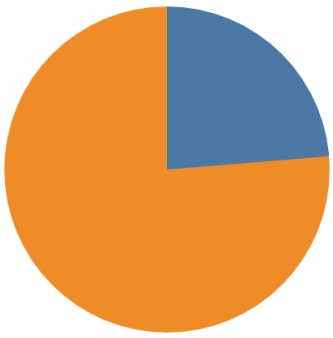
02.13

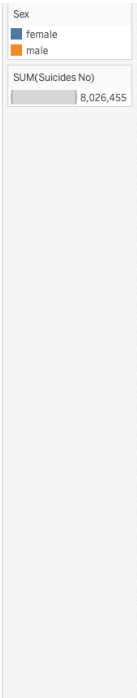
dt03 - Stacked Bar Chart - Suicides by Gender



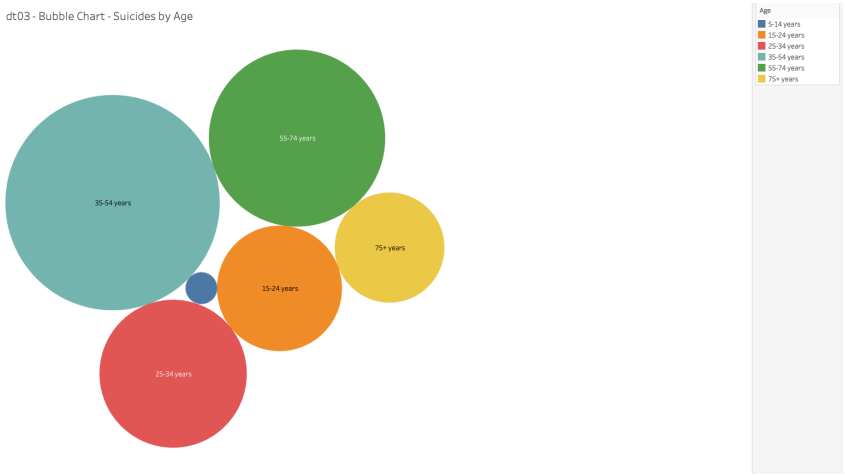
02.14

dt03 - Pie Chart - Suicides by Gender

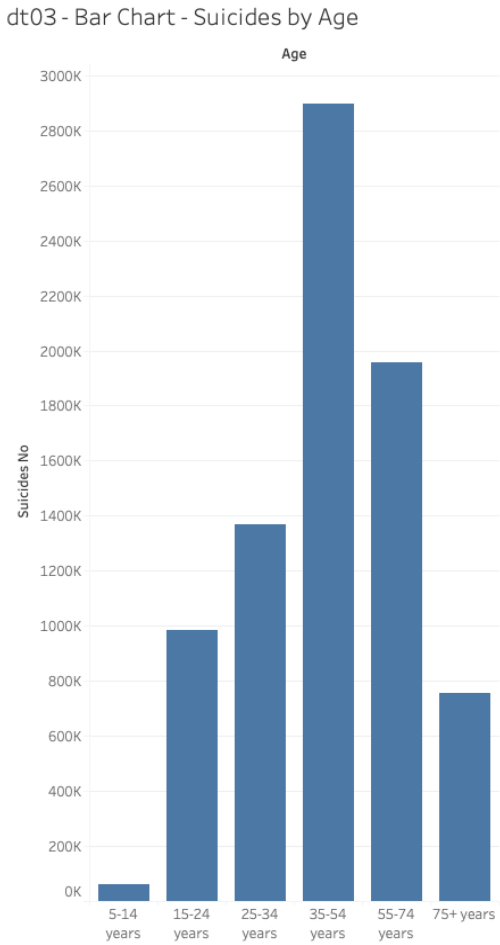




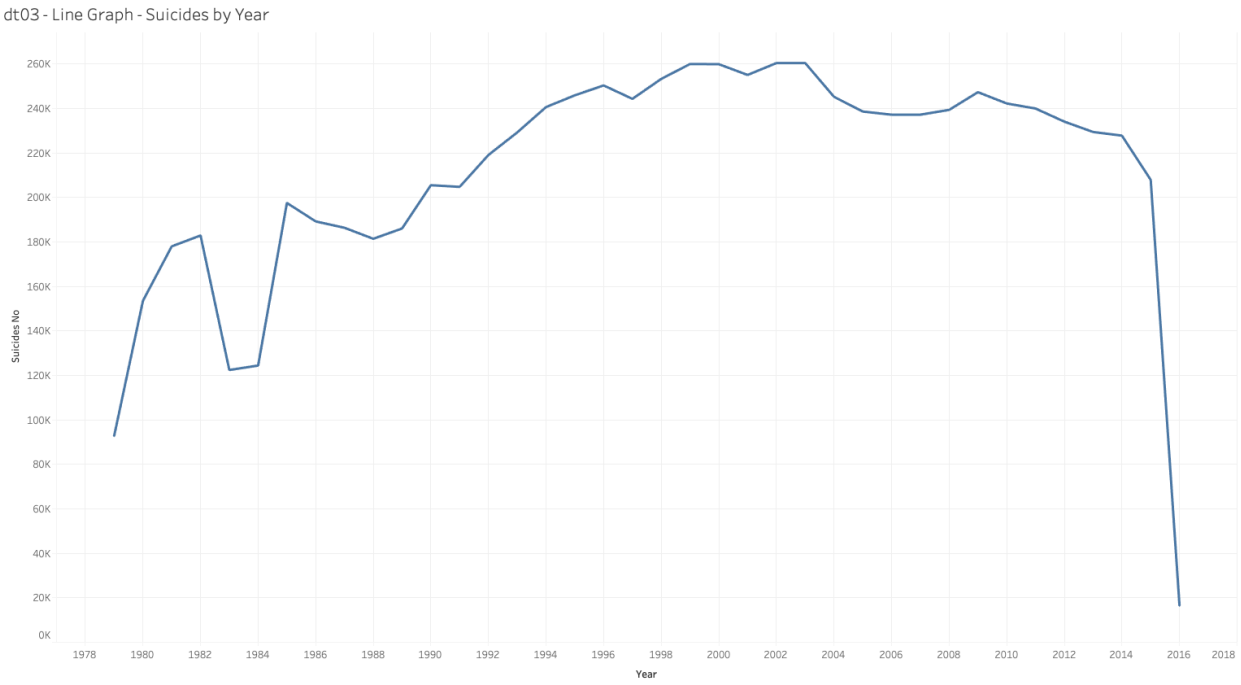
02.15


dt03 - Bubble Chart - Suicides by Age
55-74 years
35-54 years
75+ years
15-24 years
25-34 years
Age
5-14 years
15-24 years
25-34 years
35-54 years
55-74 years
75+ years

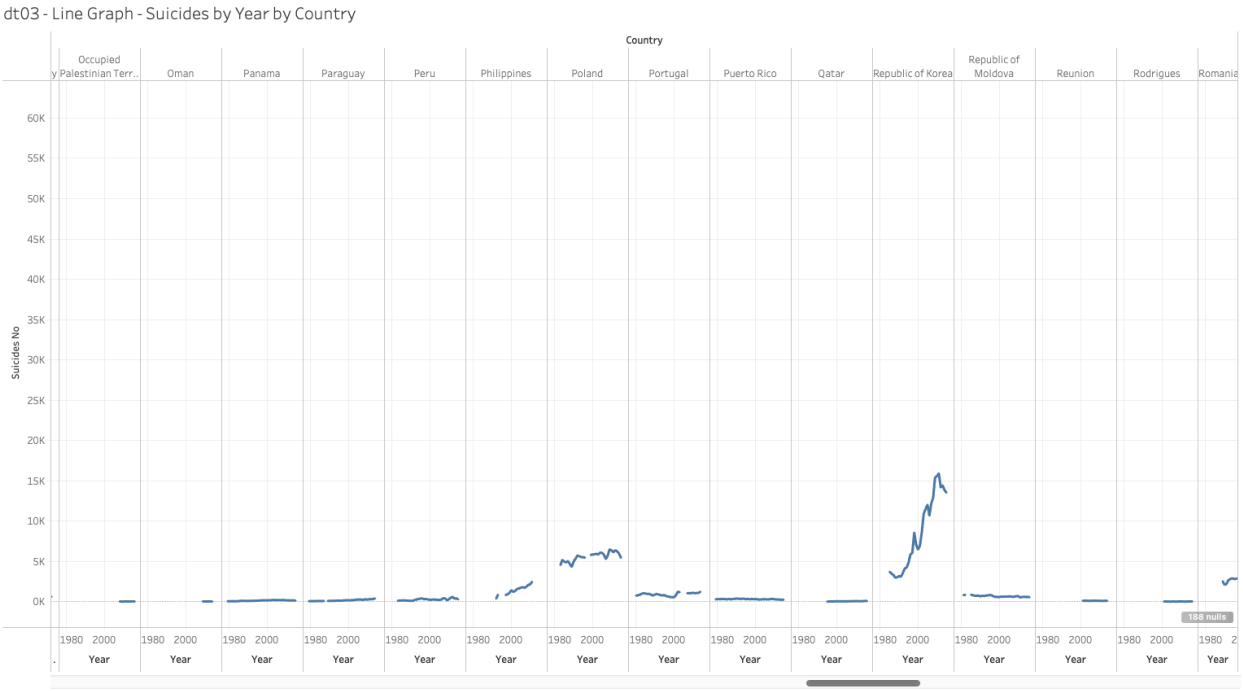
02.16


dt03 - Bar Chart - Suicides by Age
Age
Suicides No
3000K
2800K
2600K
2400K
2200K
2000K
1800K
1600K
1400K
1200K
1000K
800K
600K
400K
200K
0K
5-14 years
15-24 years
25-34 years
35-54 years
55-74 years
75+ years

02.17


dt03 - Line Graph - Suicides by Year
Suicides No
260K
240K
220K
200K
180K
160K
140K
120K
100K
80K
60K
40K
20K
0K
1978 1980 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2018
Year

02.18


dt03 - Line Graph - Suicides by Year by Country
Country
Occupied Palestinian Terr..
Oman
Panama
Paraguay
Peru
Philippines
Poland
Portugal
Puerto Rico
Qatar
Republic of Korea
Republic of Moldova
Reunion
Rodrigues
Romania
Suicides No
60K
55K
50K
45K
40K
35K
30K
25K
20K
15K
10K
5K
0K
1980 2000
Year
188 nulls

03.01

```
# 10.02.01.01
# read csv
# assign variable
# dt04

dt04_combined_____________00 = pd.read_csv('combined_processed_data.csv')
```

```
# 10.02.01.02
# read csv
# assign variable
# dt05

dt05_stress_general_______00 = pd.read_csv('Degree_of_Stress__General_Life__13_years_old_and_over__20240719092712.csv')
```

```
# 10.02.01.03
# read csv
# assign variable
# dt06

dt06_stress_home__________00 = pd.read_csv('Degree_of_Stress__Home_Life__13_years_old_and_over__20240719092914.csv')
```

```
# 10.02.01.04
# read csv
# assign variable
# dt07

dt07_stress_school________00 = pd.read_csv('Degree_of_Stress__School_Life__13_years_old_and_over__20240719092757.csv', encoding = 'unicode_escape')
```

```
# 10.02.01.05
# read csv
# assign variable
# dt08

dt08_stress_work__________00 = pd.read_csv('Degree_of_Stress__Work_Life__13_years_old_and_over__20240719092838.csv')
```

```
# 10.02.01.06
# read csv
# assign variable
# dt09

dt09_drinking_19__________00 = pd.read_csv('Drinking__19_years_old_and_over__20240719093320.csv')
```

03.02

```
# 10.02.01.07
# read csv
# assign variable
# dt10

dt10_drinking_20__________00 = pd.read_csv('Drinking__20_years_old_and_over__20240719093241.csv')
```

```
# 10.02.01.08
# read csv
# assign variable
# dt11

dt11_drinking_manage_19____00 = pd.read_csv('Drinking_and_Health_Management__19_years_old_and_over__20240719093528.csv')
```

```
# 10.02.01.09
# read csv
# assign variable
# dt12

dt12_drinking_manage_20____00 = pd.read_csv('Drinking_and_Health_Management__20_years_old_and_over__20240719093453.csv')
```

```
# 10.02.01.09
# read csv
# assign variable
# dt13

dt13_suicide_impulse______00 = pd.read_csv('Impulse_to_Commit_Suicide_and_Reasons__13_years_old_and_over__20240719092337.csv')
```

```
# 10.02.01.10
# read csv
# assign variable
# dt14

dt14_suicide_impulse______00 = pd.read_csv('Impulse_to_Commit_Suicide_and_Reasons__13_years_old_and_over__20240719092337.csv')
```

```
# 10.02.01.11
# read csv
# assign variable
# dt15

dt15_suicide_reason_______00 = pd.read_csv('Reason_and_Attempt_to_Think_Suicide_by_General_Feature_of_older_persons_Over_65_Years_Old__20240719092517.csv', encoding =
'unicode_escape')
```

03.03

```
# 10.02.01.12
# read csv
# assign variable
# dt16

dt16_suicide_reason_____00 = pd.read_csv('Reason_and_Attempt_to_Think_Suicide_by_General_Feature_of_older_persons_Over_65_Years_Old__20240719092517.csv', encoding = 'unicode_escape')
```

```
# 10.02.01.13
# read csv
# assign variable
# dt17

dt17_smoke_drink_19_____00 = pd.read_csv('Smoking_and_Drinking__19_years_old_and_over__20240719093138.csv')
```

```
# 10.02.01.14
# read csv
# assign variable
# dt18

dt18_smoke_drink_20_____00 = pd.read_csv('Smoking_and_Drinking__20_years_old_and_over__20240719093056.csv')
```

```
# 10.02.01.15
# read csv
# assign variable
# dt19

dt19_ph_categories_____00 = pd.read_csv('ph_categories_index.csv')
```

```
# 10.02.01.16
# read csv
# assign variable
# dt20

'''dt20_ph_analysis_____00 = pd.read_excel('ph_Pornhub Analysis year by year.xlsx')'''
```

```
"dt20_ph_analysis_____00 = pd.read_excel('ph_Pornhub Analysis year by year.xlsx')"
```

```
# 10.02.01.17
# read csv
# assign variable
# dt21

dt21_ph_videos_____00 = pd.read_csv('ph_videos.csv')
```

03.04

```
# 10.02.52.01
# read csv
# assign variable
# dt23

dt23_408_03_____________00 = pd.read_csv('408_DT_40803_N0003_20240801134720.csv')
```

```
# 10.02.52.02
# read csv
# assign variable
# dt24

dt24_408_04_____________00 = pd.read_csv('408_DT_40803_N0004_20240801134840.csv')
```

```
# 10.02.52.03
# read csv
# assign variable
# dt25

dt25_index_eco_sent_____00 = pd.read_csv('Economic_Sentiment_Index_20240801135039.csv')
```

```
# 10.02.52.04
# read csv
# assign variable
# dt26

dt26_sadness____________00 = pd.read_csv('Feeling_sad_or_hopeless_20240801134129.csv')
```

```
# 10.02.52.05
# read csv
# assign variable
# dt27

dt27_happiness_01_______00 = pd.read_csv('happiness.csv')
```

```
# 10.02.52.06
# read csv
# assign variable
# dt28

dt28_happiness_2015_____00 = pd.read_csv('index_happiness_2015.csv')
```

03.05

```
# 10.02.52.07
# read csv
# assign variable
# dt29

dt29_happiness_2016__________00 = pd.read_csv('index_happiness_2016.csv')
```

```
# 10.02.52.08
# read csv
# assign variable
# dt30

dt30_happiness_2017__________00 = pd.read_csv('index_happiness_2017.csv')
```

```
# 10.02.52.09
# read csv
# assign variable
# dt31

dt31_happiness_2018__________00 = pd.read_csv('index_happiness_2018.csv')
```

```
# 10.02.52.10
# read csv
# assign variable
# dt32

dt32_happiness_2019__________00 = pd.read_csv('index_happiness_2019.csv')
```

```
# 10.02.52.11
# read csv
# assign variable
# dt33

dt33_pop_income_____________00 = pd.read_csv('Population_by_income_level_20240801133314.csv')
```

```
# 10.02.52.12
# read csv
# assign variable
# dt34

dt34_stress_perc____________00 = pd.read_csv('Preceived_stress_20240801134258.csv')
```

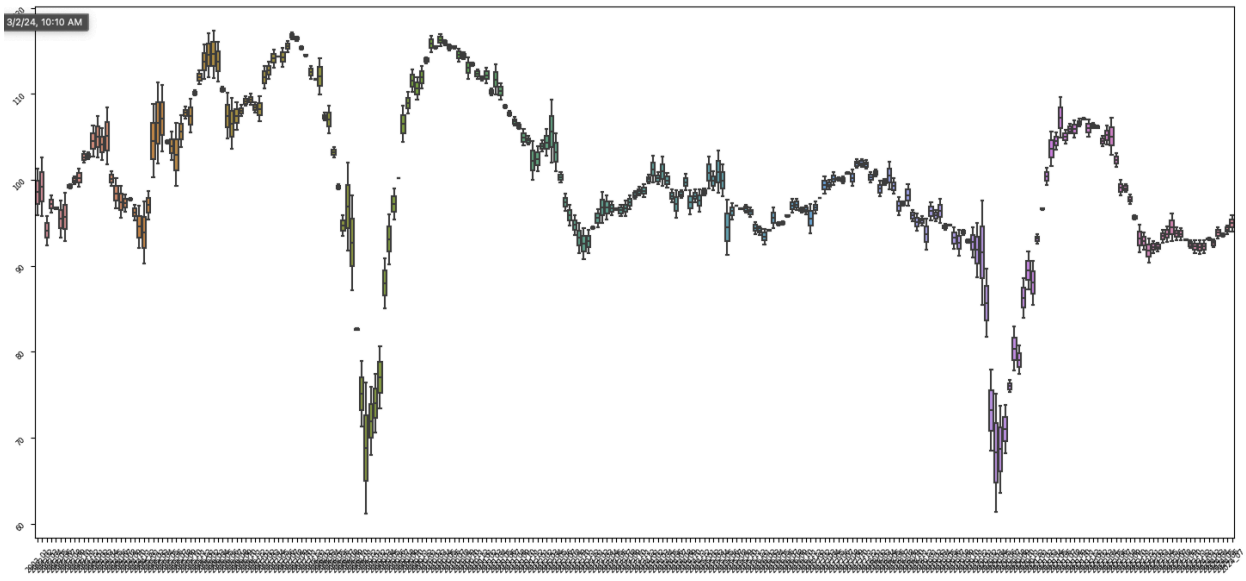
03.06

```
# 10.02.52.13
# read csv
# assign variable
# dt35

dt35_suic_reason_____________00 = pd.read_csv('Reason_and_Attempt_to_Think_Suicide_by_General_Feature_of_older_persons_Over_65_Years_Old__20240801134427.csv', encoding='unicode_escape')
```

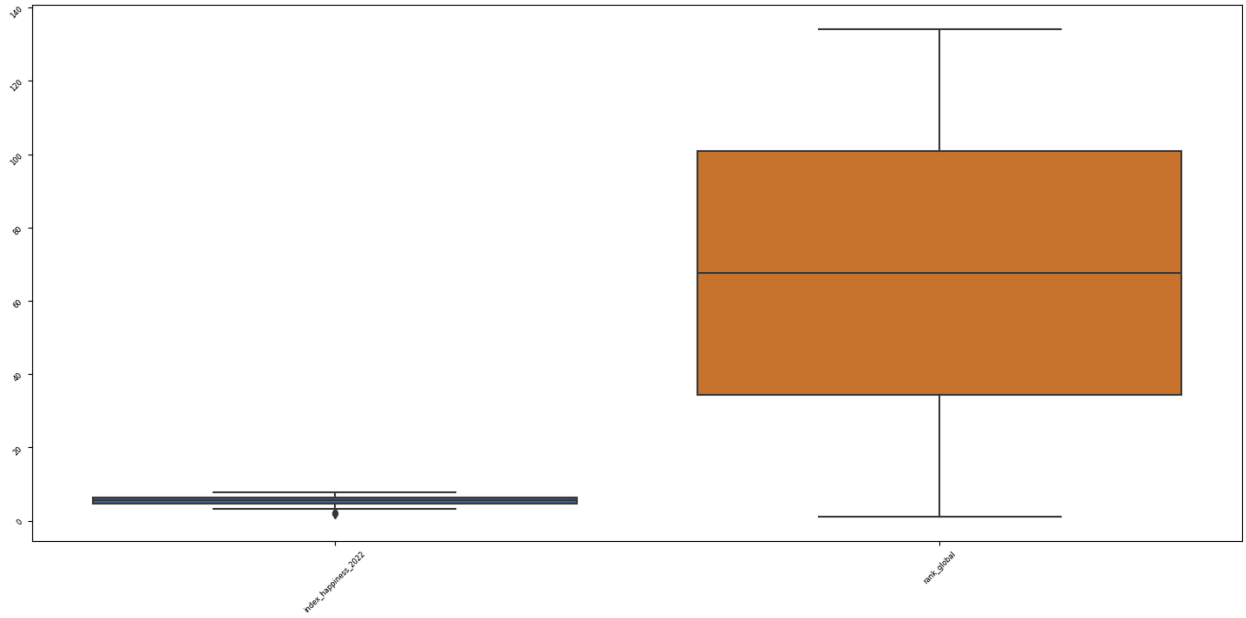
```
# 10.02.52.14
# read csv
# assign variable
# dt36

dt36_depr_symptom_____________00 = pd.read_csv('Symptom_of_Depression_by_General_Feature_of_older_persons_Over_65_Years_Old__20240801135254.csv', encoding='unicode_escape')
```

```
# 10.02.52.15
# read csv
# assign variable
# dt37

dt37_happiness_world_________00 = pd.read_csv('World Happiness Report.csv')
```

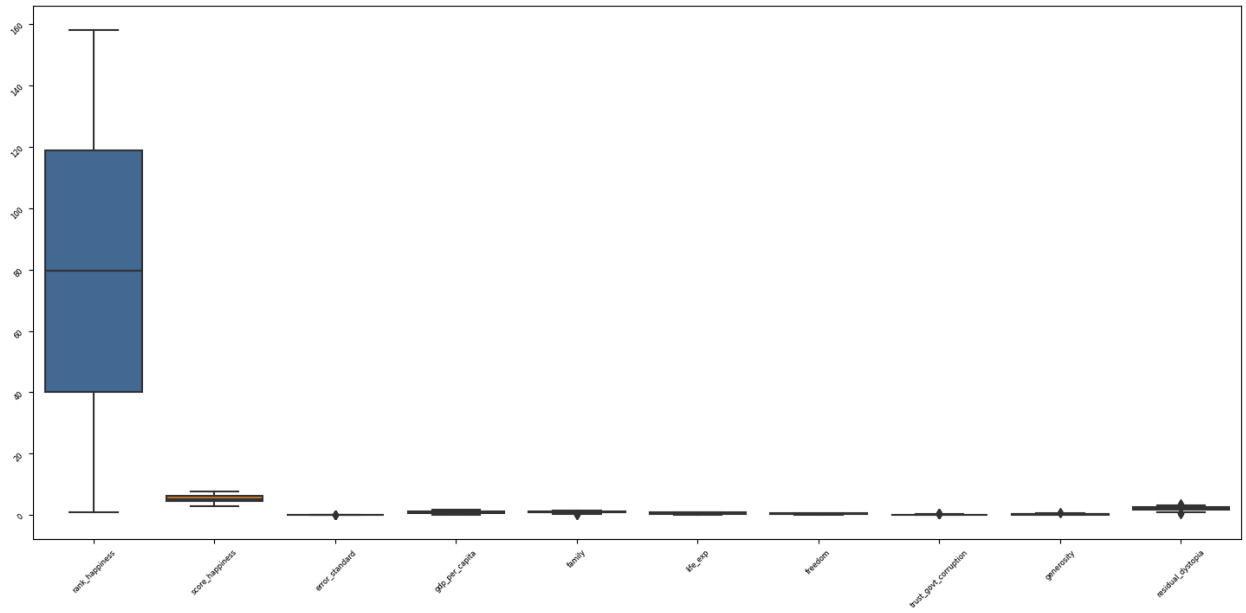
03.07



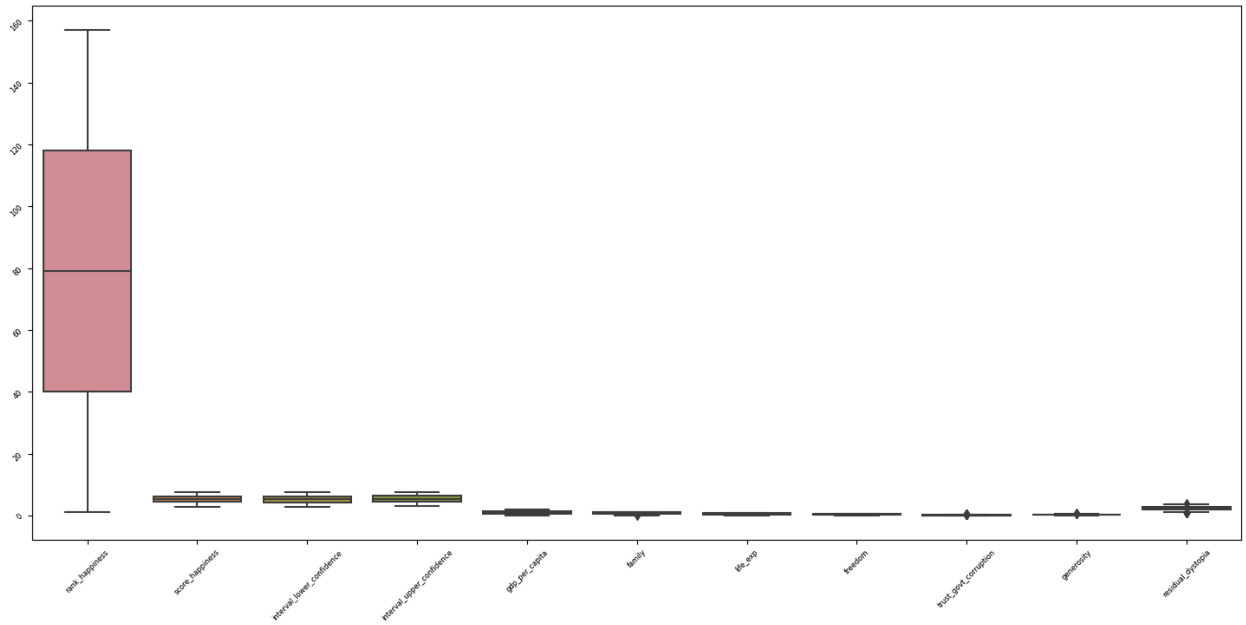
03.08


140
120
100
80
60
40
20
0
index_happiness_2022
rank_global

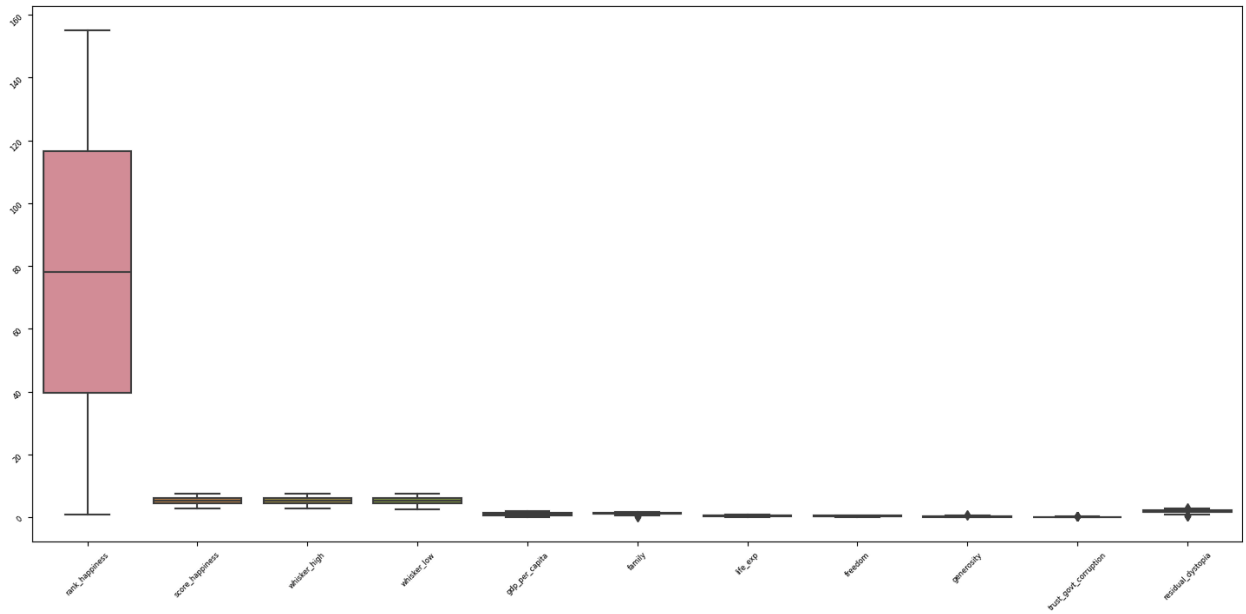
03.09


160
140
120
100
80
60
40
20
0
rank_happiness
score_happiness
error_standard
gdp_per_capita
family
life_exp
freedom
trust_govt_corruption
generosity
residual_dystopia

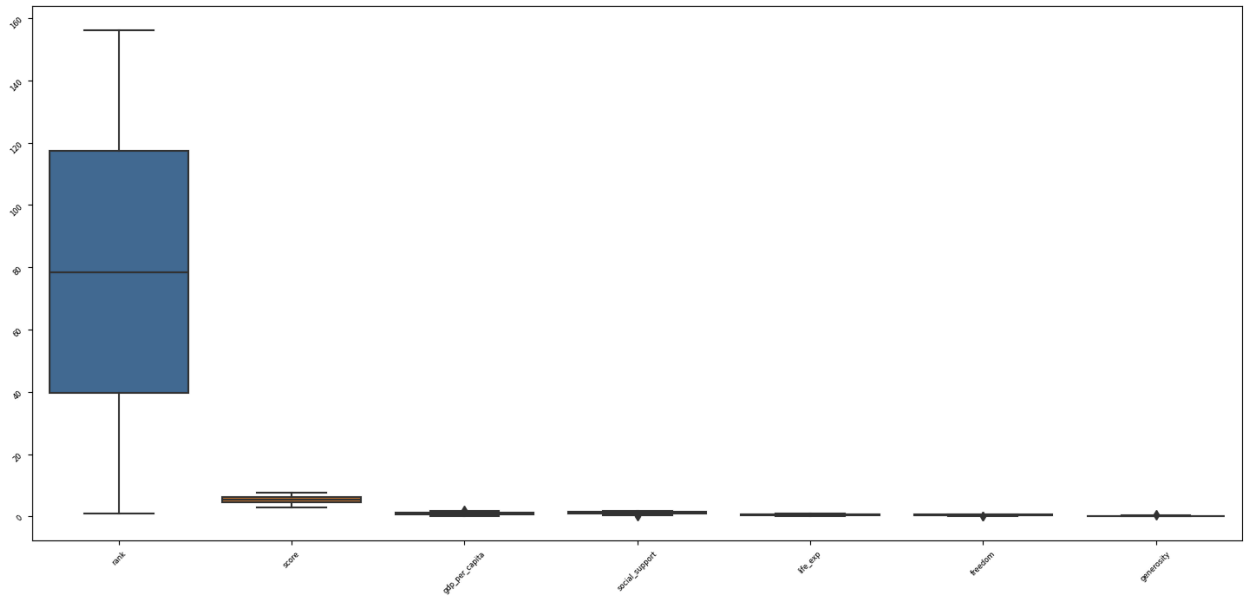
03.10


160
140
120
100
80
60
40
20
0
rank_happiness
score_happiness
interval_lower_confidence
interval_upper_confidence
gdp_per_capita
family
life_exp
freedom
trust_govt_corruption
generosity
residual_dystopia

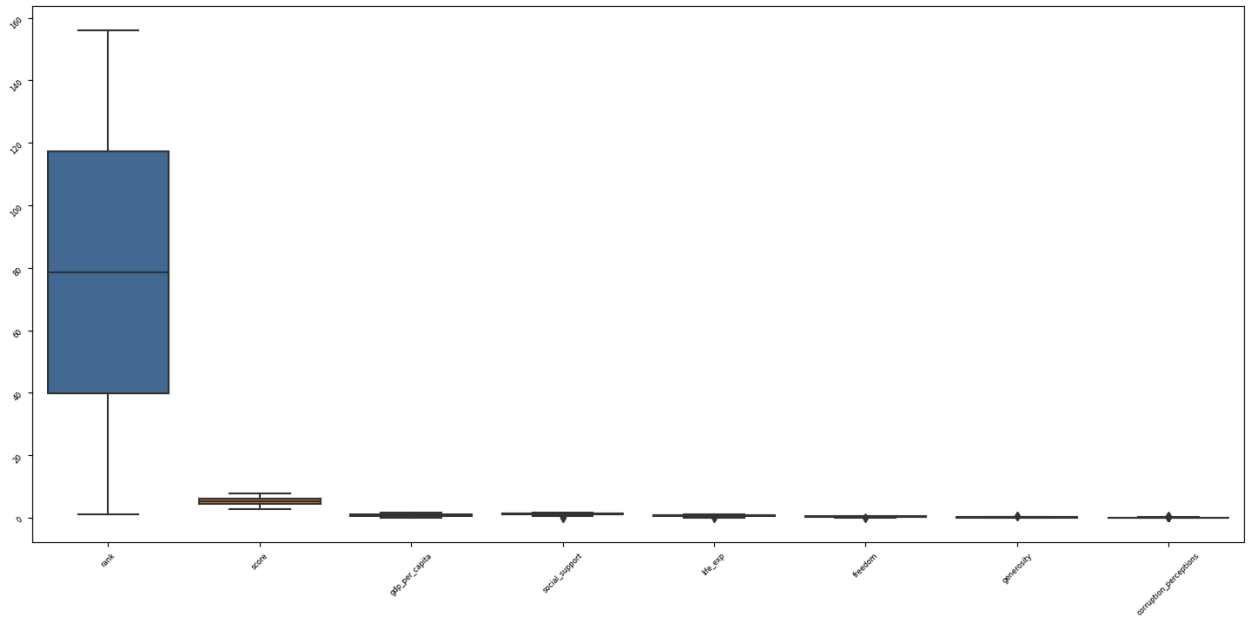
03.11


160
140
120
100
80
60
40
20
0
rank_happiness
score_happiness
whisker_high
whisker_low
gdp_per_capita
family
life_exp
freedom
generosity
trust_govt_corruption
residual_dystopia

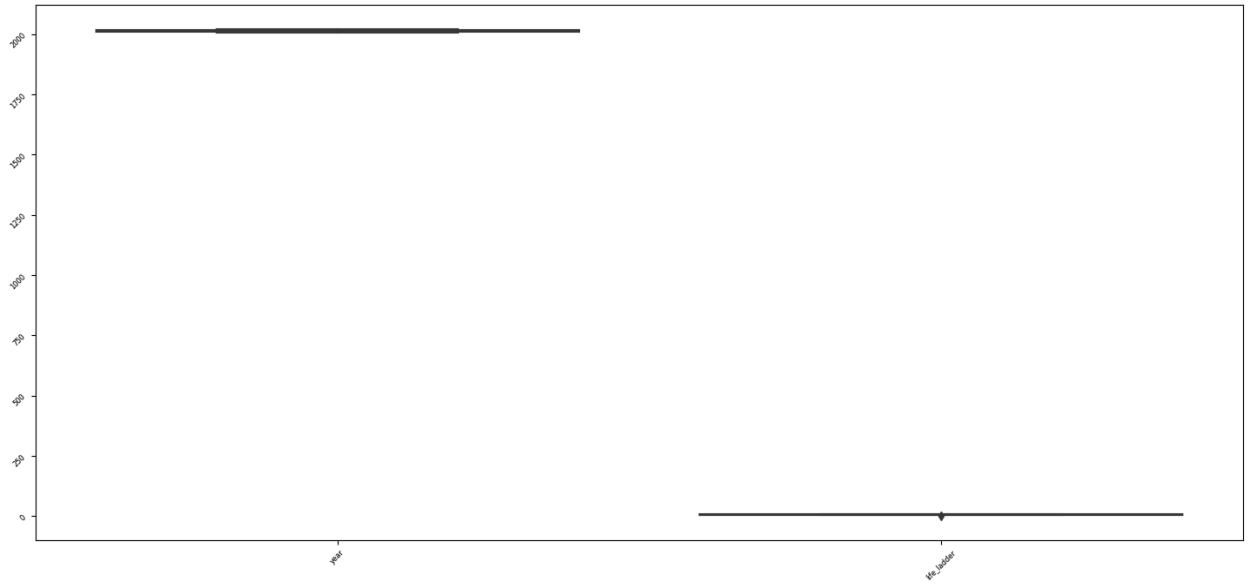
03.12


160
140
120
100
80
60
40
20
0
rank
score
gdp_per_capita
social_support
life_exp
freedom
generosity

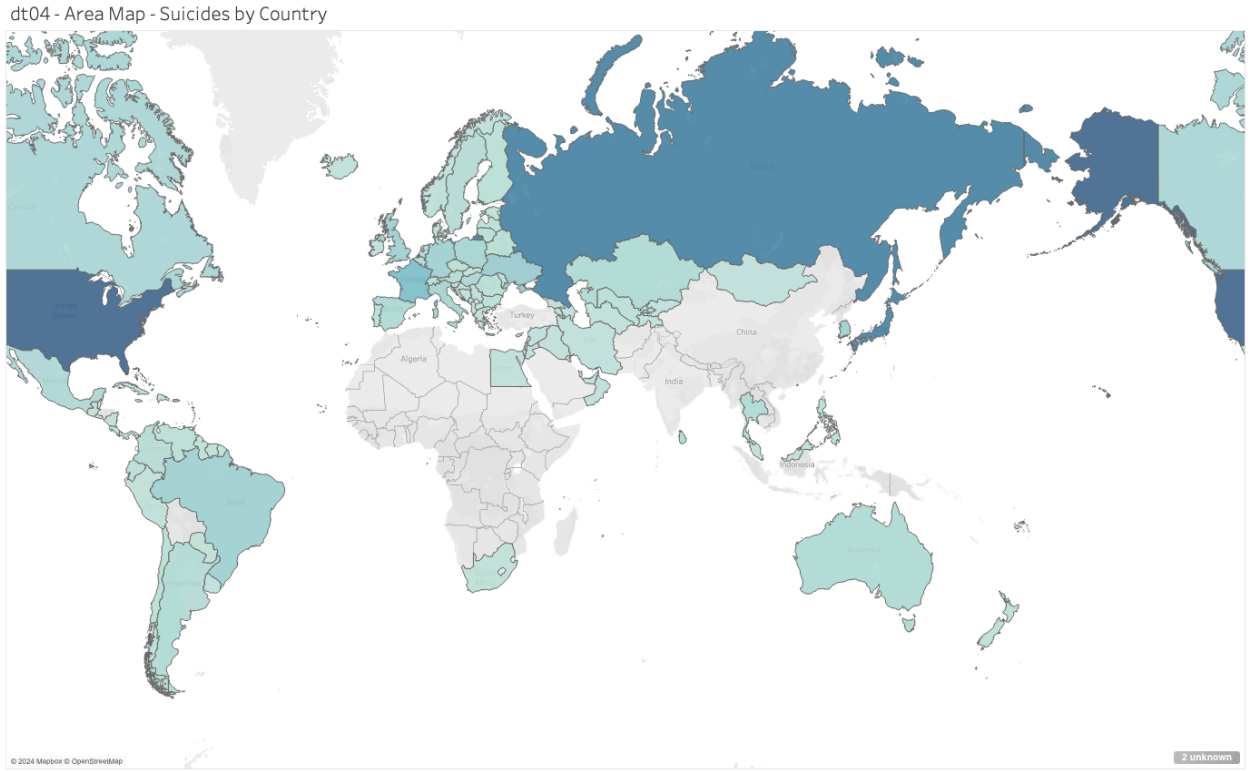
03.13


160
140
120
100
80
60
40
20
0
rank
score
gdp_per_capita
social_support
life_exp
freedom
generosity
corruption_perceptions

03.14

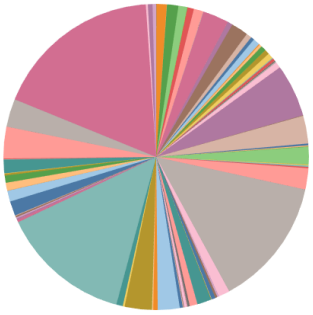

2000
1750
1500
1250
1000
750
500
250
0
year
life_ladder

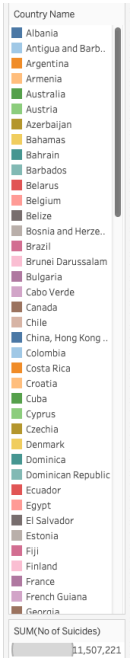
03.15


dt04 - Area Map - Suicides by Country
Russia
Canada
United States
Turkey
China
Algeria
India
Indonesia
Brazil
Australia
© 2024 Mapbox © OpenStreetMap
2 unknown

03.16

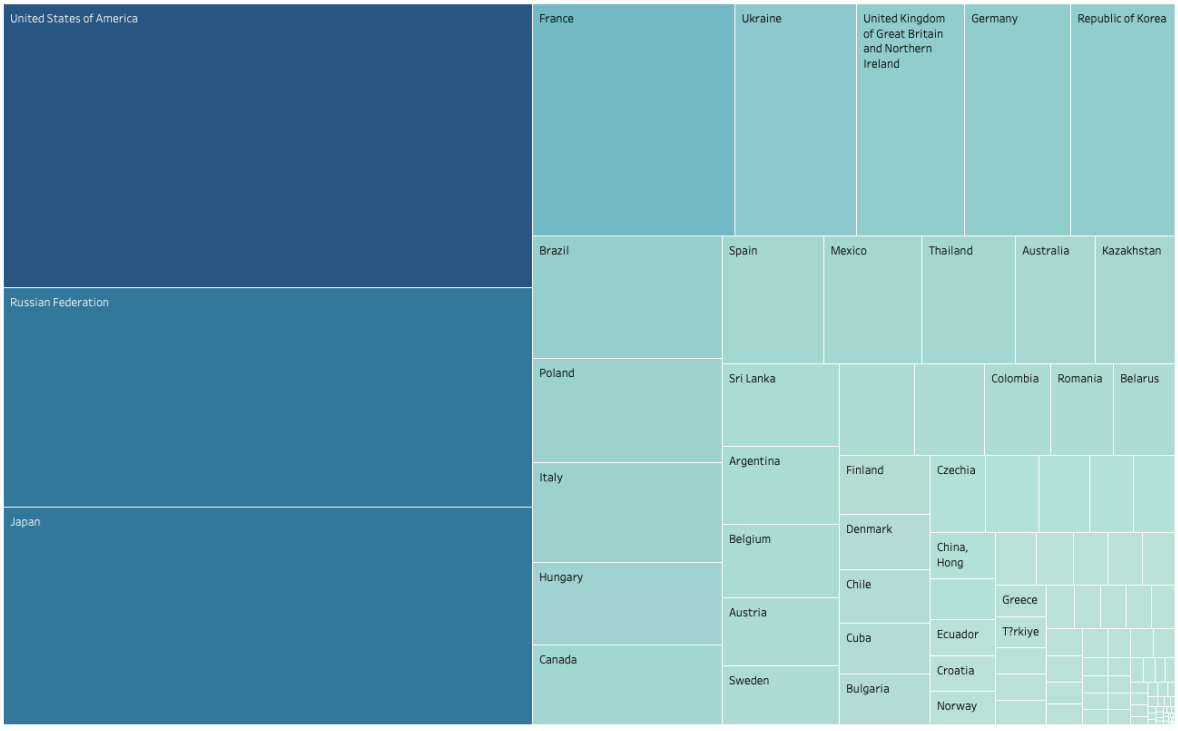
dt04 - Pie Chart - Suicides by Country



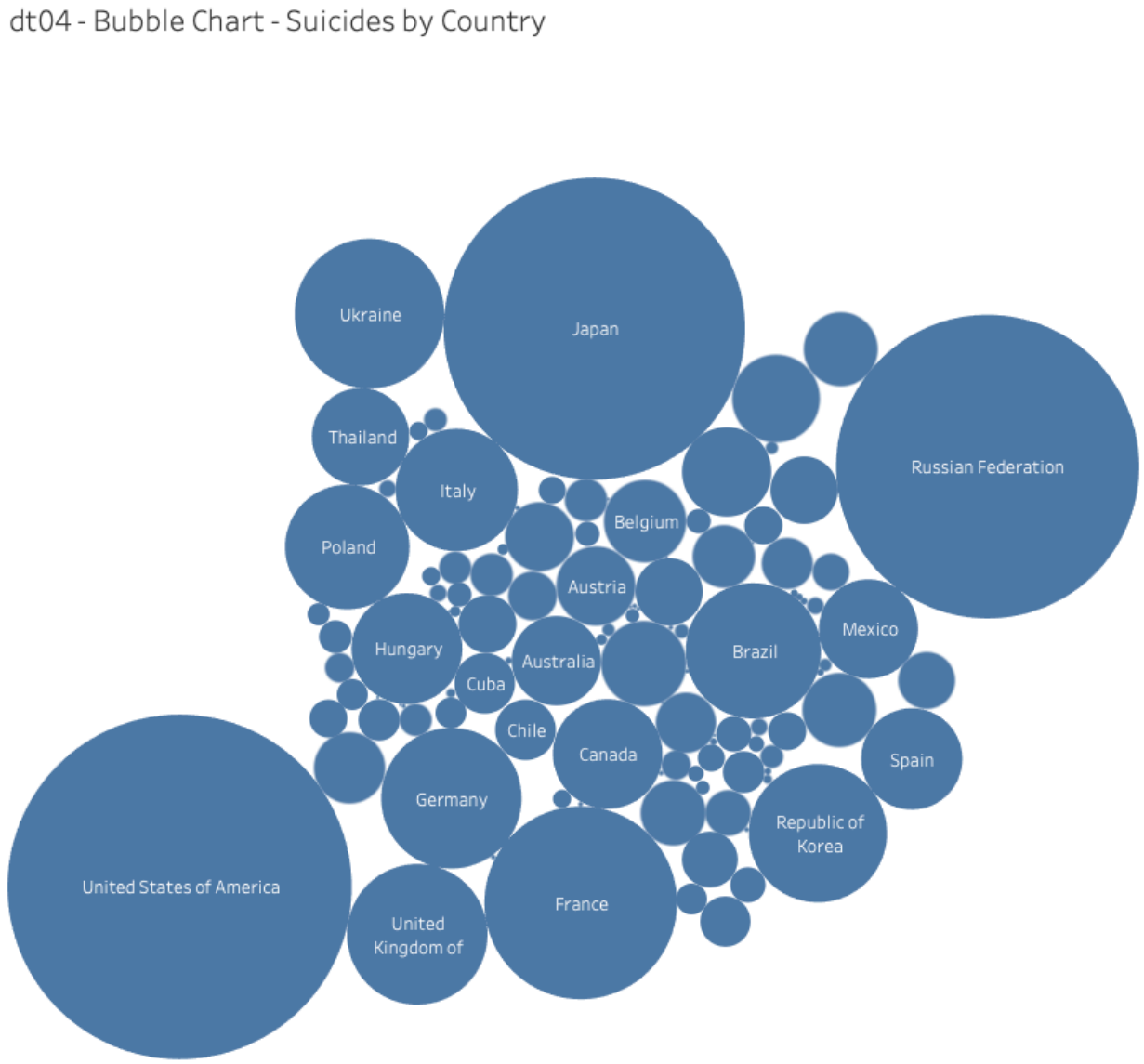


03.17

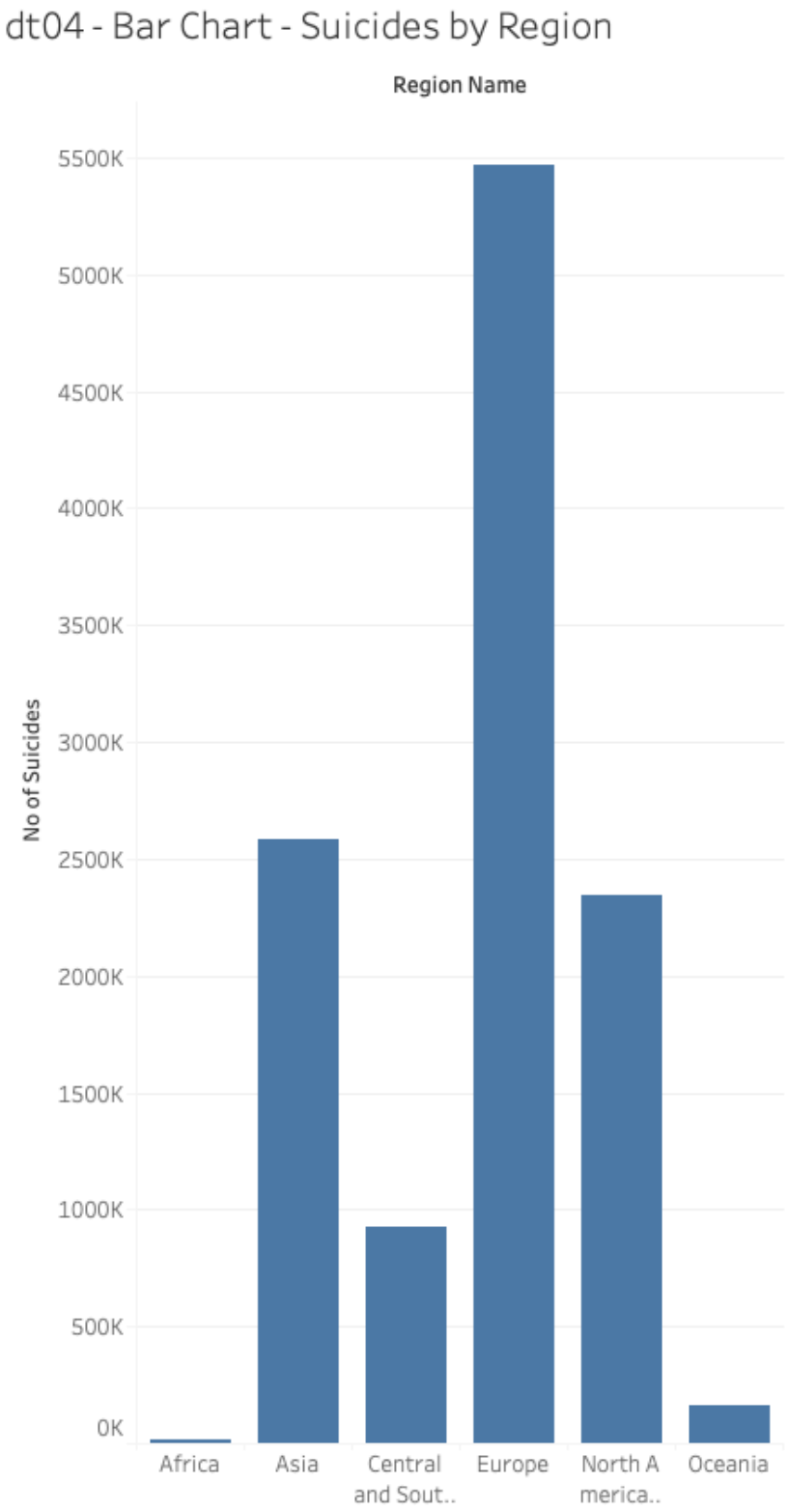
dt04 - Tree Map - Suicides by Country



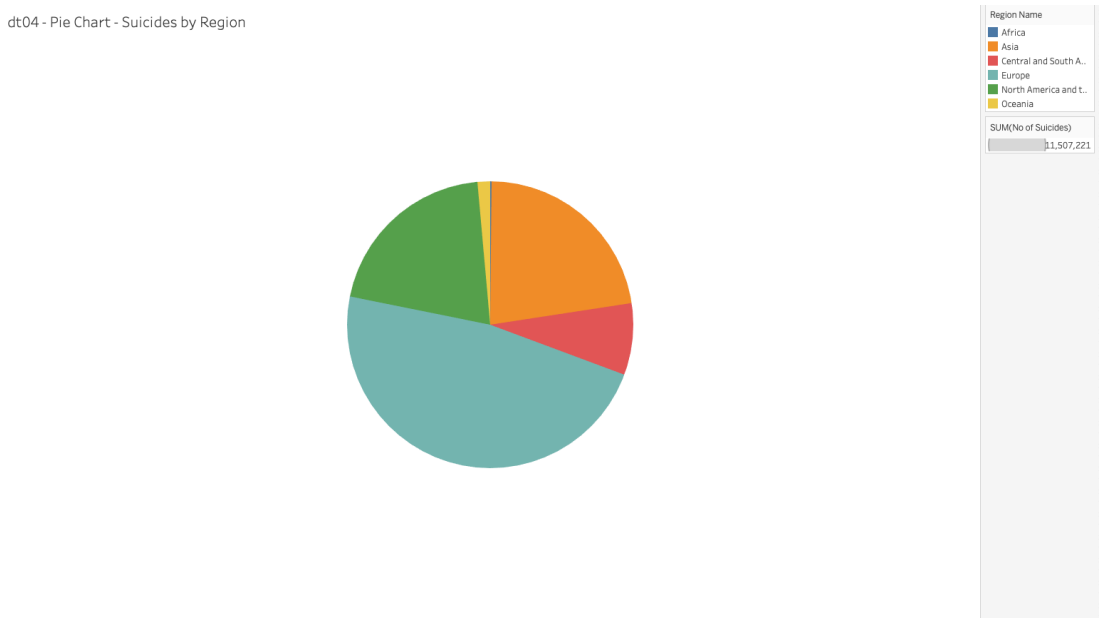
03.18


dt04 - Bubble Chart - Suicides by Country
Ukraine
Japan
Thailand
Russian Federation
Italy
Belgium
Poland
Austria
Hungary
Australia
Brazil
Mexico
Cuba
Chile
Canada
Spain
Germany
Republic of Korea
United States of America
United Kingdom of
France

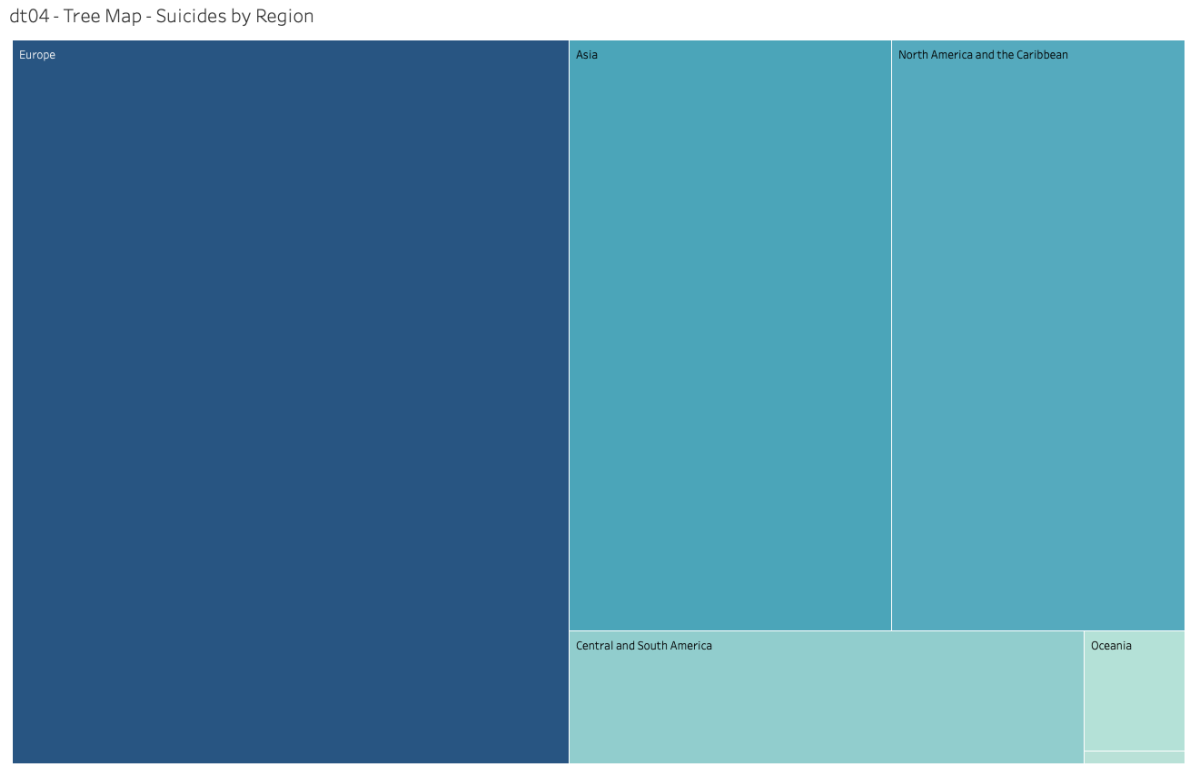
03.19


dt04 - Bar Chart - Suicides by Region
Region Name
No of Suicides
5500K
5000K
4500K
4000K
3500K
3000K
2500K
2000K
1500K
1000K
500K
0K
Africa
Asia
Central and Sout..
Europe
North America..
Oceania

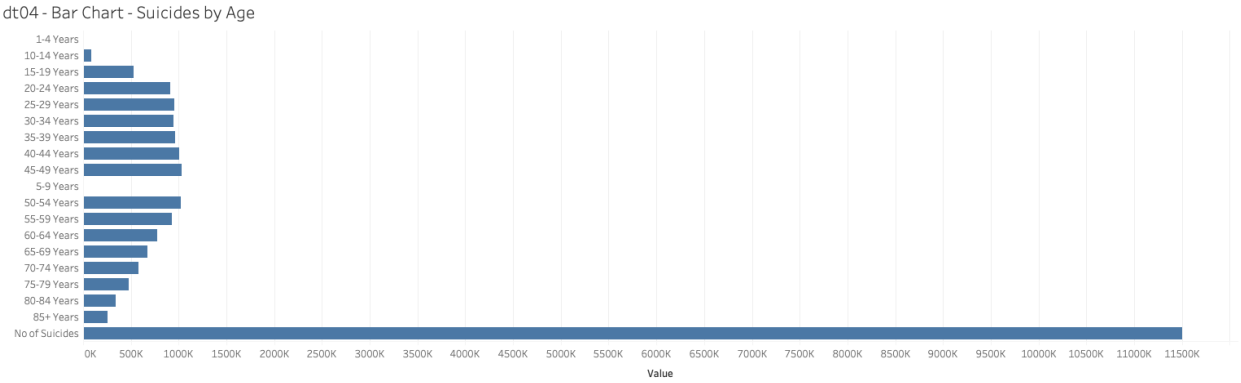
03.20


dt04 - Pie Chart - Suicides by Region
Region Name
Africa
Asia
Central and South A..
Europe
North America and t..
Oceania
SUM(No of Suicides)
11,507,221

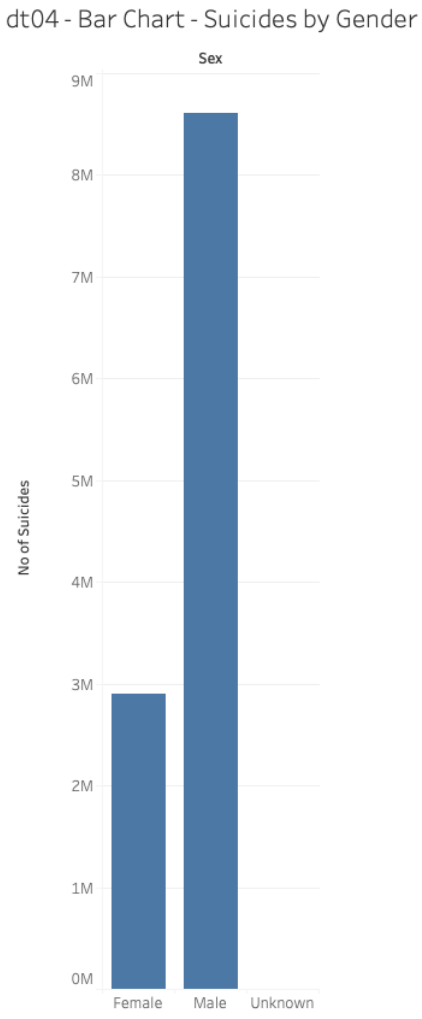
03.21


dt04 - Tree Map - Suicides by Region
Europe
Asia
North America and the Caribbean
Central and South America
Oceania

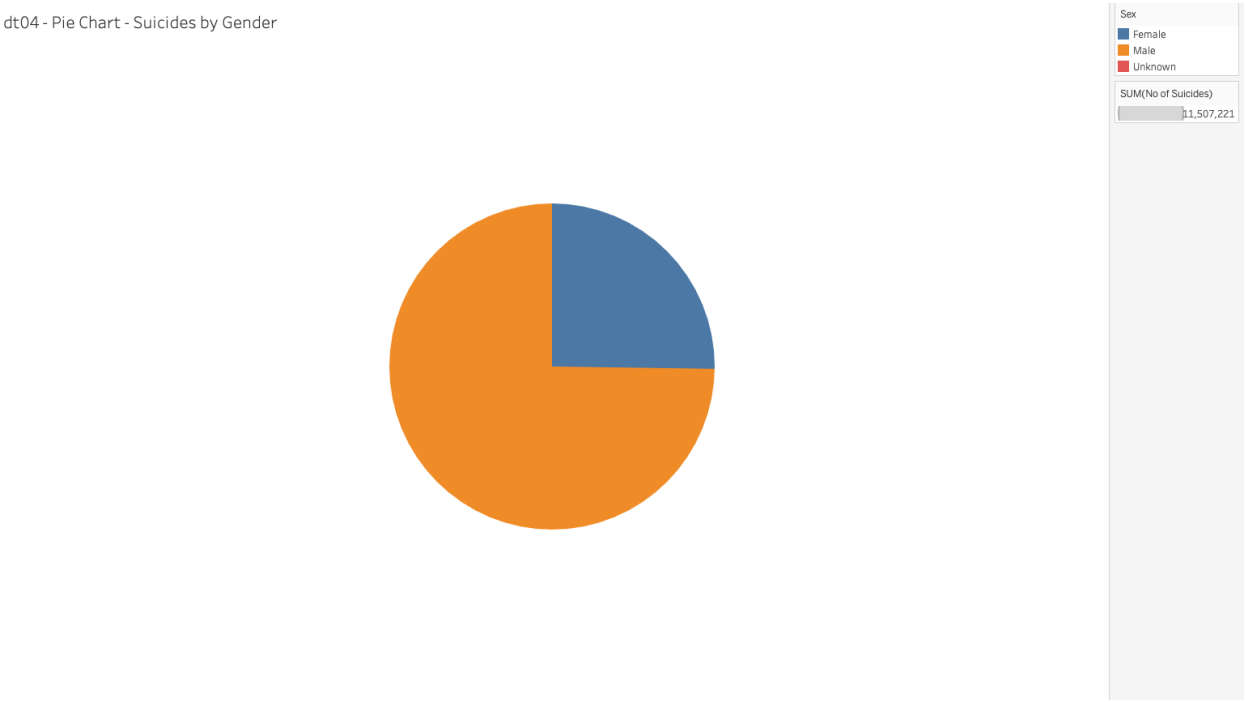
03.22


dt04 - Bar Chart - Suicides by Age
1-4 Years
10-14 Years
15-19 Years
20-24 Years
25-29 Years
30-34 Years
35-39 Years
40-44 Years
45-49 Years
5-9 Years
50-54 Years
55-59 Years
60-64 Years
65-69 Years
70-74 Years
75-79 Years
80-84 Years
85+ Years
No of Suicides
0K 500K 1000K 1500K 2000K 2500K 3000K 3500K 4000K 4500K 5000K 5500K 6000K 6500K 7000K 7500K 8000K 8500K 9000K 9500K 10000K 10500K 11000K 11500K
Value

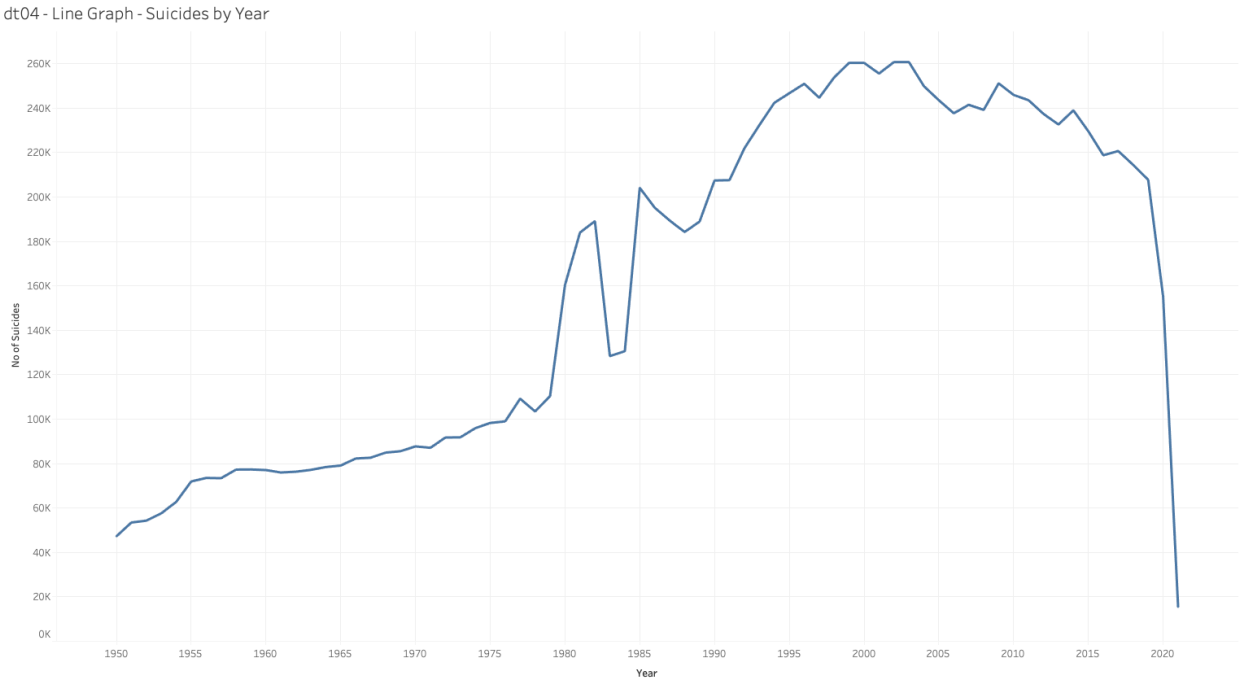
03.23


dt04 - Bar Chart - Suicides by Gender
Sex
9M
8M
7M
6M
5M
4M
3M
2M
1M
0M
No of Suicides
Female
Male
Unknown

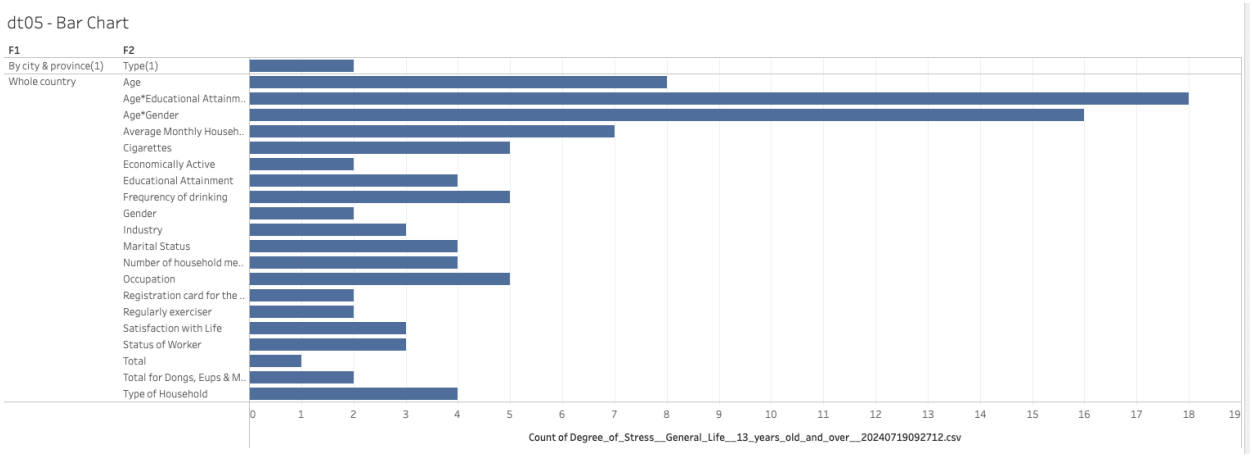
03.24


dt04 - Pie Chart - Suicides by Gender
Sex
Female
Male
Unknown
SUM(No of Suicides)
11,507,221

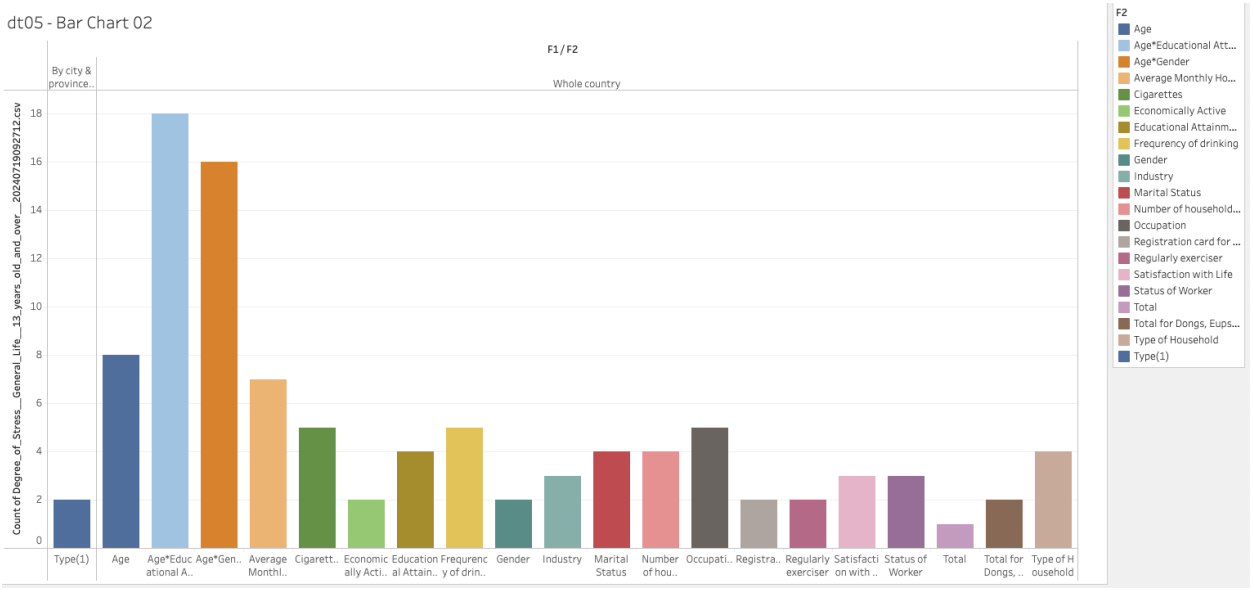
03.25


dt04 - Line Graph - Suicides by Year
No of Suicides
260K
240K
220K
200K
180K
160K
140K
120K
100K
80K
60K
40K
20K
0K
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Year

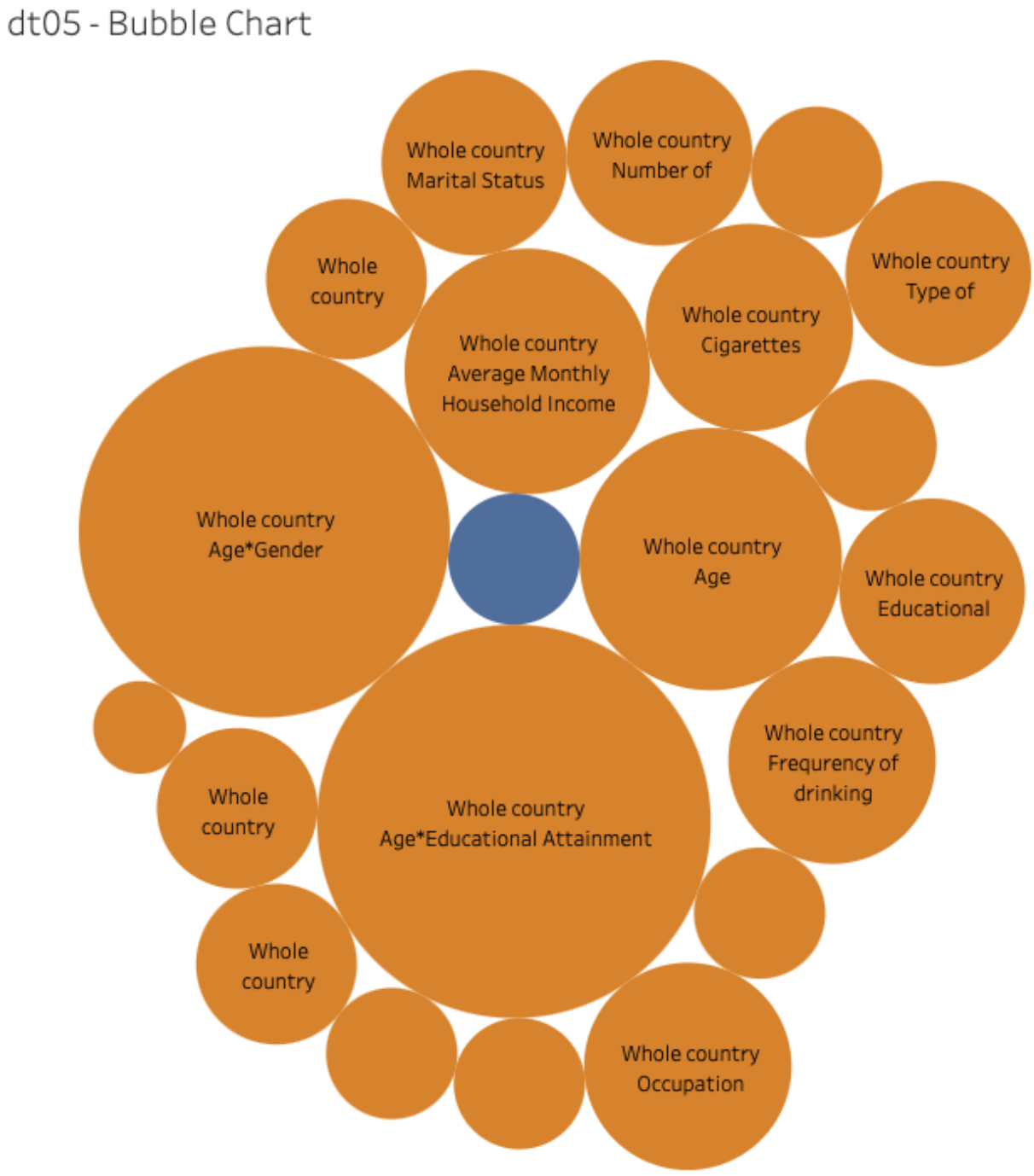
03.26


dt05 - Bar Chart
F1
F2
By city & province(1)
Whole country
Type(1)
Age
Age*Educational Attainm..
Age*Gender
Average Monthly Househ..
Cigarettes
Economically Active
Educational Attainment
Frequrency of drinking
Gender
Industry
Marital Status
Number of household me..
Occupation
Registration card for the ..
Regularly exerciser
Satisfaction with Life
Status of Worker
Total
Total for Dongs, Eups & M..
Type of Household
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Count of Degree_of_Stress__General_Life__13_years_old_and_over__20240719092712.csv

03.27

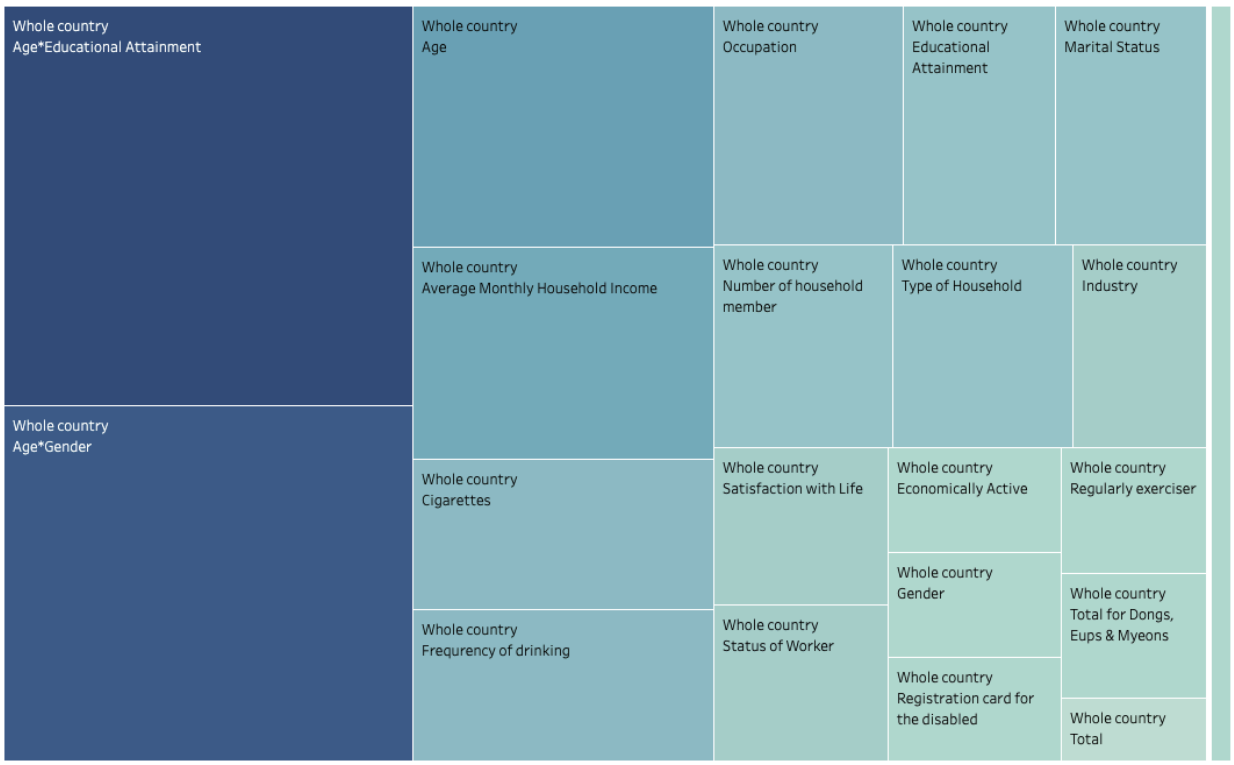

dt05 - Bar Chart 02
F1 / F2
By city & province..
Whole country
Count of Degree_of_Stress__General_Life__13_years_old_and_over__20240719092712.csv
0 2 4 6 8 10 12 14 16 18
Type(1)
Age
Age*Educational A..
Age*Gen..
Average Monthl..
Cigarett..
Economically Acti..
Education al Attain..
Frequenc y of drin..
Gender
Industry
Marital Status
Number of hou..
Occupati..
Registra..
Regularly exerciser
Satisfacti on with ..
Status of Worker
Total
Total for Dongs, ..
Type of H ousehold
F2
Age
Age*Educational Att...
Age*Gender
Average Monthly Ho...
Cigarettes
Economically Active
Educational Attainm...
Frequrency of drinking
Gender
Industry
Marital Status
Number of household...
Occupation
Registration card for ...
Regularly exerciser
Satisfaction with Life
Status of Worker
Total
Total for Dongs, Eups...
Type of Household
Type(1)

03.28


dt05 - Bubble Chart
Whole country Marital Status
Whole country Number of
Whole country
Whole country Type of
Whole country Cigarettes
Whole country Average Monthly Household Income
Whole country Age*Gender
Whole country Age
Whole country Educational
Whole country Frequency of drinking
Whole country
Whole country Age*Educational Attainment
Whole country
Whole country Occupation

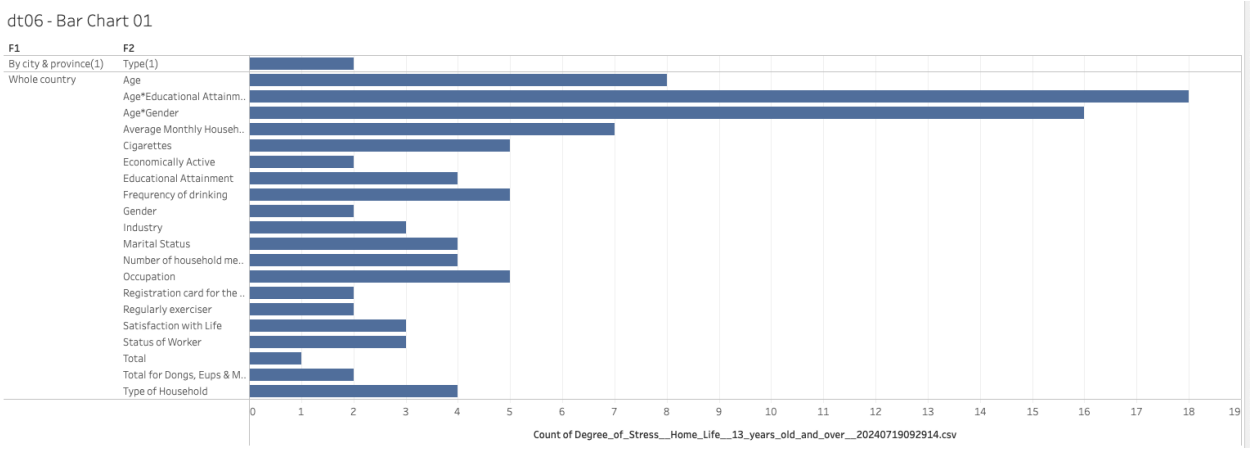
03.29

dt05 - Tree Map

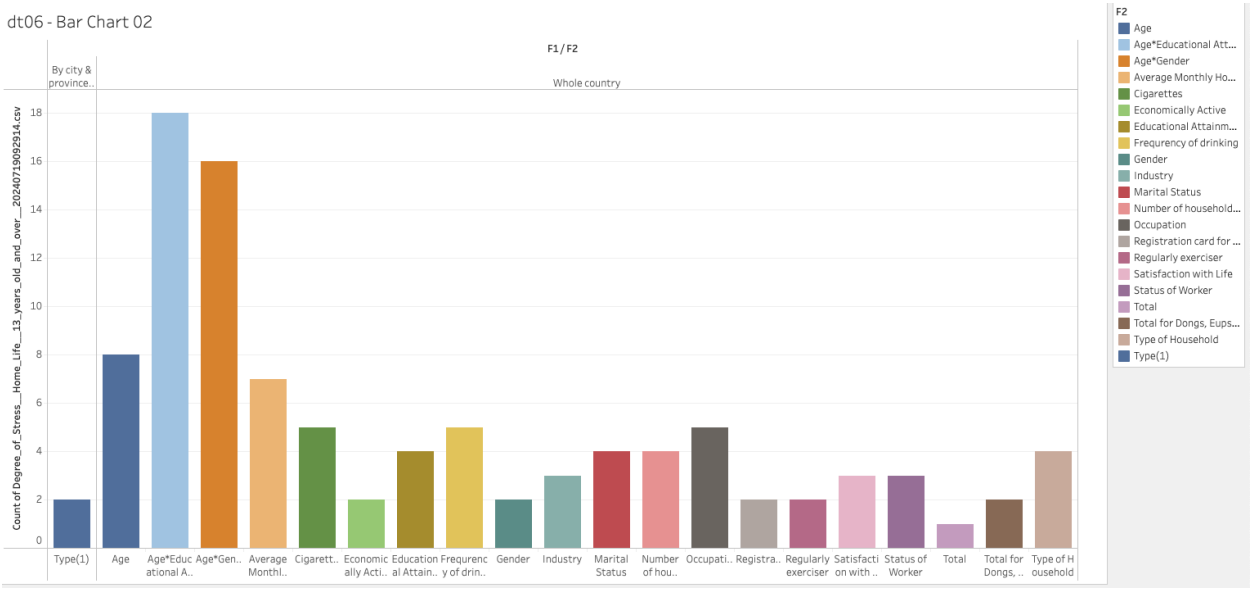


03.30

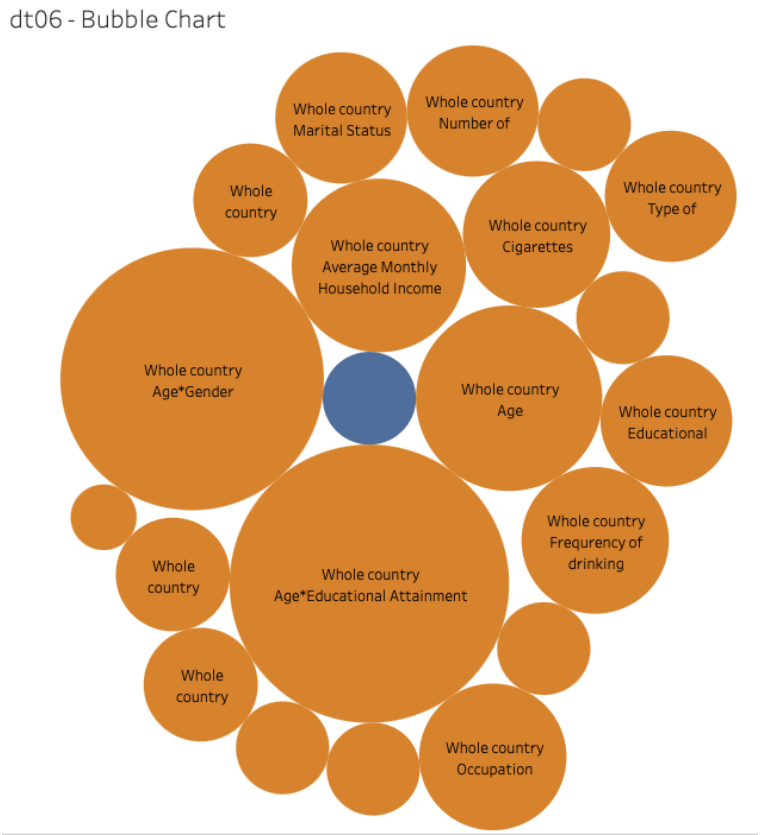
dt06 - Bar Chart 01



03.31

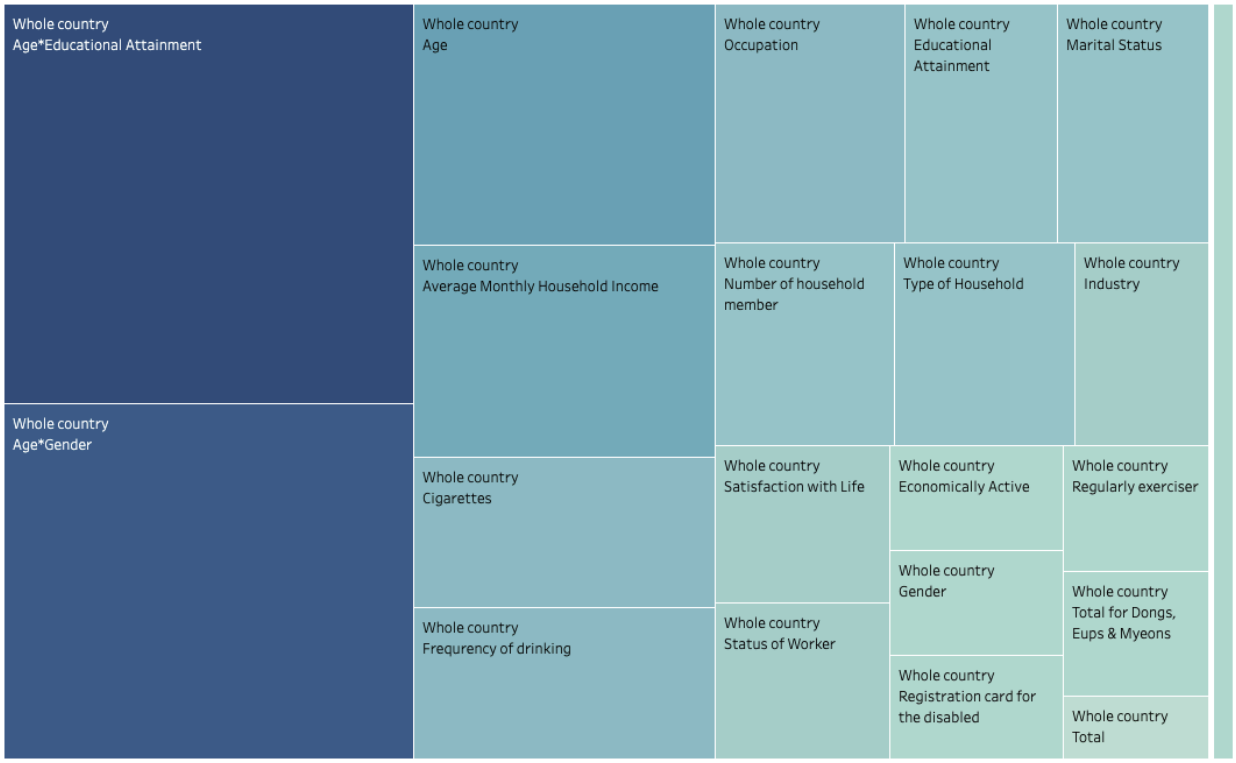

dt06 - Bar Chart 02
F1 / F2
By city & province..
Whole country
Count of Degree_of_Stress__Home_Life__13_years_old_and_over__20240719092914.csv
0
2
4
6
8
10
12
14
16
18
Type(1)
Age
Age*Educational A..
Age*Gen..
Average Monthl..
Cigarett..
Economically Acti..
Educational Attain..
Frequency of drin..
Gender
Industry
Marital Status
Number of hou..
Occupati..
Registra..
Regularly exerciser
Satisfaction with ..
Status of Worker
Total
Total for Dongs, ..
Type of Household
F2
Age
Age*Educational Att...
Age*Gender
Average Monthly Ho...
Cigarettes
Economically Active
Educational Attainm...
Frequency of drinking
Gender
Industry
Marital Status
Number of household...
Occupation
Registration card for ...
Regularly exerciser
Satisfaction with Life
Status of Worker
Total
Total for Dongs, Eups...
Type of Household
Type(1)

03.32


dt06 - Bubble Chart
Whole country Marital Status
Whole country Number of
Whole country
Whole country Type of
Whole country Cigarettes
Whole country Average Monthly Household Income
Whole country Age*Gender
Whole country Age
Whole country Educational
Whole country Frequency of drinking
Whole country
Whole country Age*Educational Attainment
Whole country
Whole country Occupation

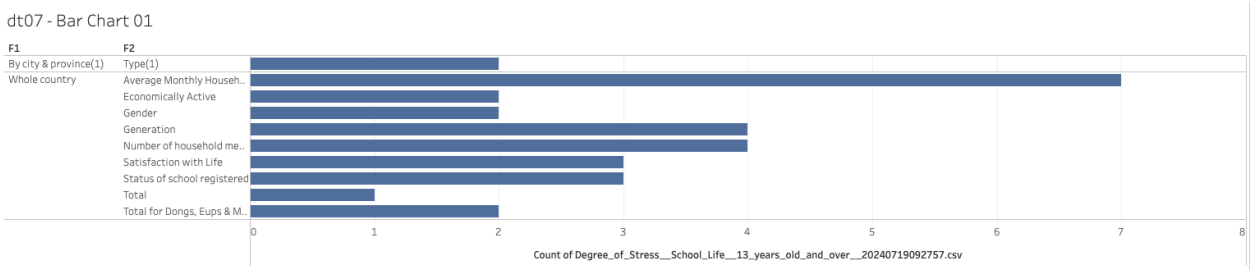
03.33

dt06 - Tree Map

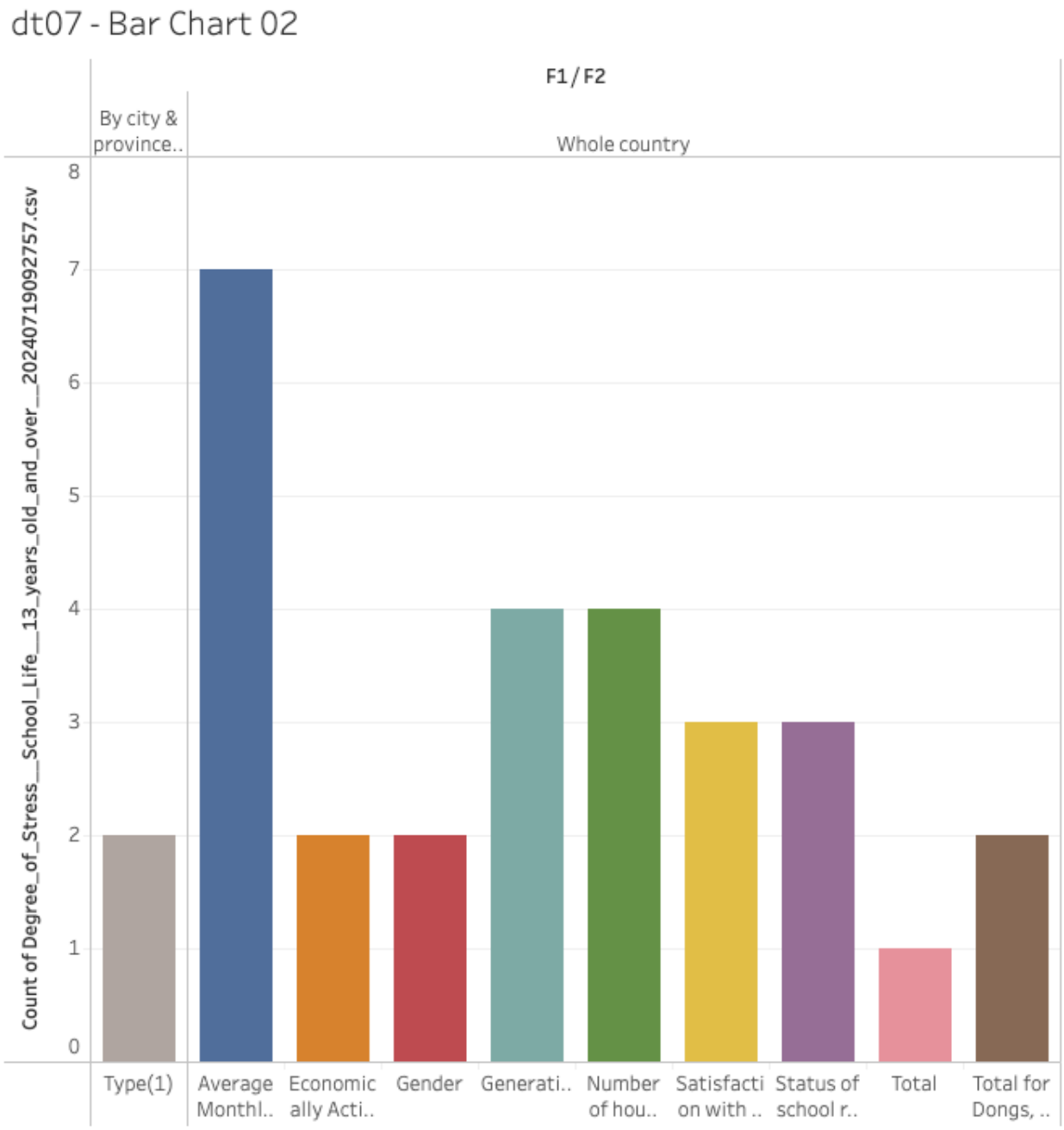


03.34

dt07 - Bar Chart 01



03.35

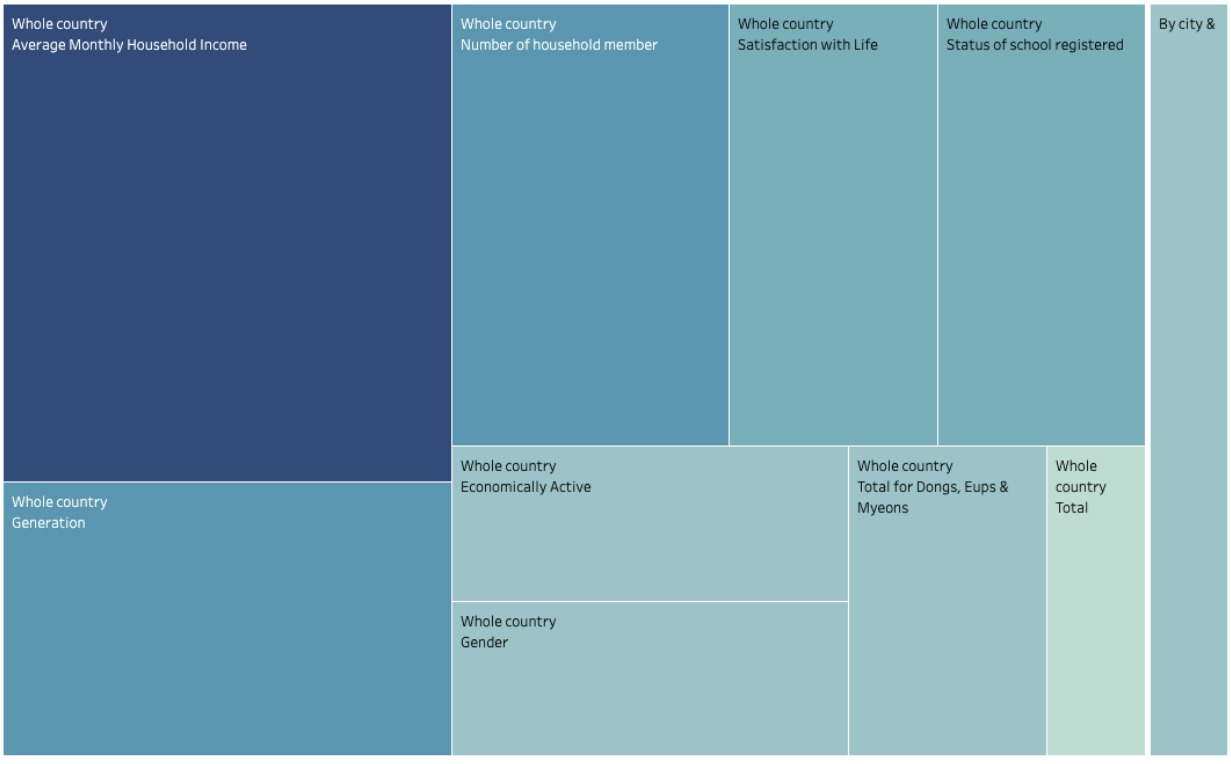

dt07 - Bar Chart 02
F1 / F2
By city & province..
Whole country
Count of Degree_of_Stress__School_Life__13_years_old_and_over__20240719092757.csv
0
1
2
3
4
5
6
7
8
Type(1)
Average Monthl..
Economic ally Acti..
Gender
Generati..
Number of hou..
Satisfacti on with ..
Status of school r..
Total
Total for Dongs, ..

03.36

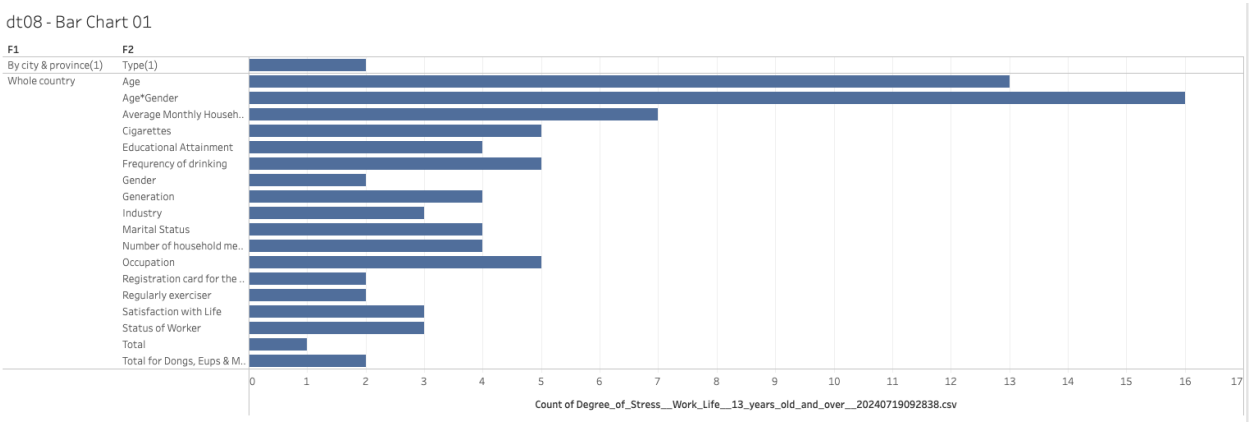
dt07 - Bubble Chart

03.37

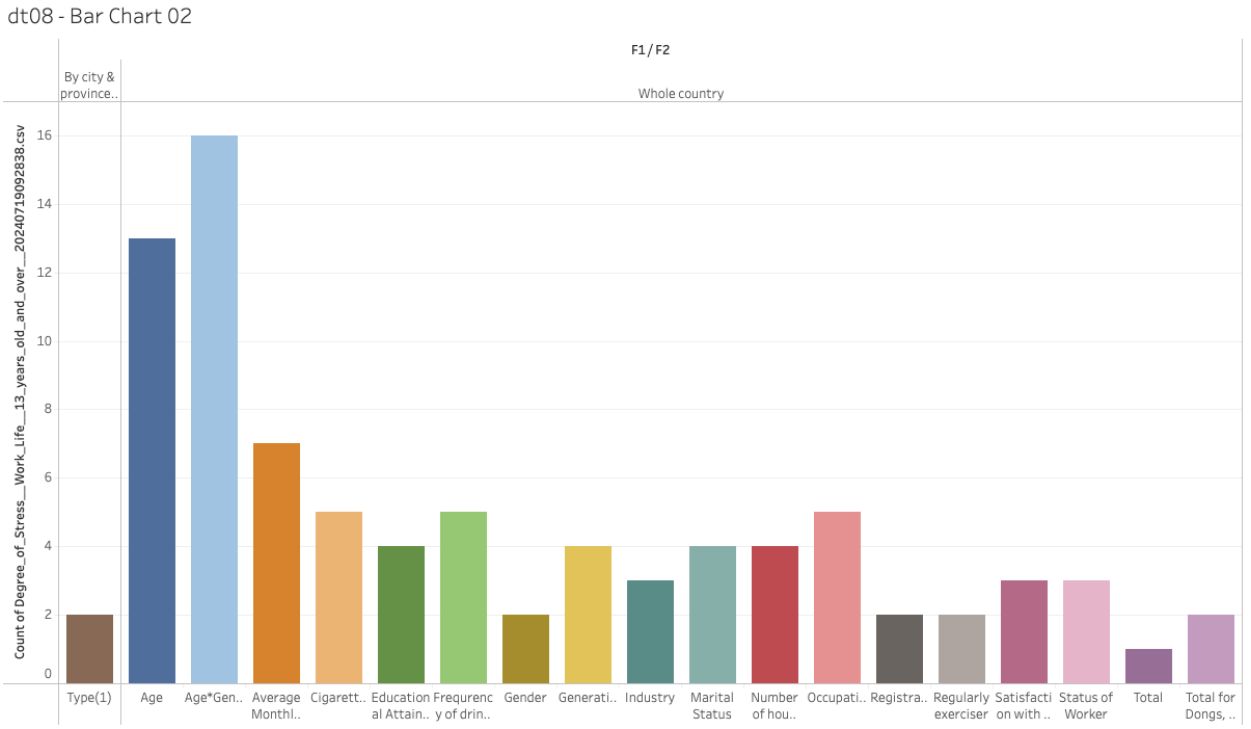
dt07 - Tree Map



03.38


dt08 - Bar Chart 01
F1
F2
By city & province(1)
Type(1)
Whole country
Age
Age*Gender
Average Monthly Househ..
Cigarettes
Educational Attainment
Frequrency of drinking
Gender
Generation
Industry
Marital Status
Number of household me..
Occupation
Registration card for the ..
Regularly exerciser
Satisfaction with Life
Status of Worker
Total
Total for Dongs, Eups & M..
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
Count of Degree_of_Stress__Work_Life__13_years_old_and_over__20240719092838.csv

03.39

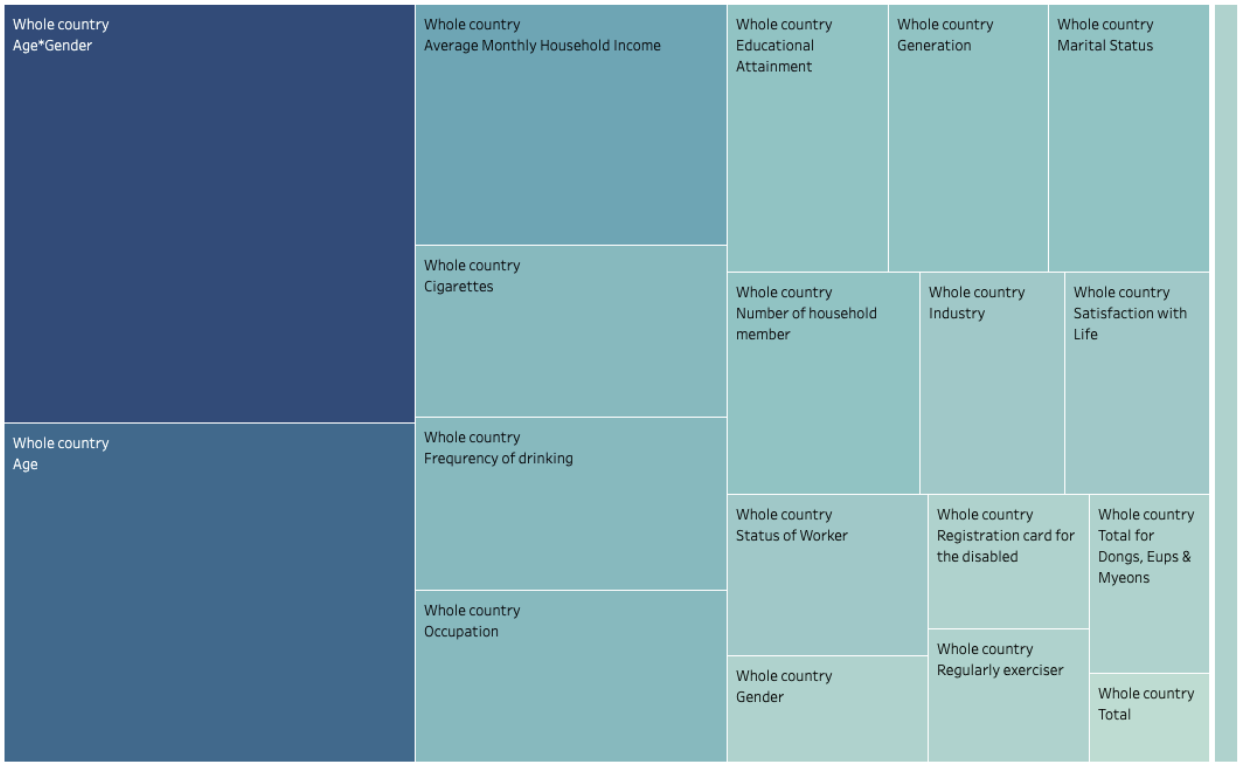

dt08 - Bar Chart 02
F1 / F2
By city & province..
Whole country
Count of Degree_of_Stress__Work_Life__13_years_old_and_over__20240719092838.csv
0 2 4 6 8 10 12 14 16
Type(1)
Age
Age*Gen..
Average Monthl..
Cigarett..
Education al Attain..
Frequenc y of drin..
Gender
Generati..
Industry
Marital Status
Number of hou..
Occupati..
Registra..
Regularly exerciser
Satisfacti on with ..
Status of Worker
Total
Total for Dongs, ..

03.40

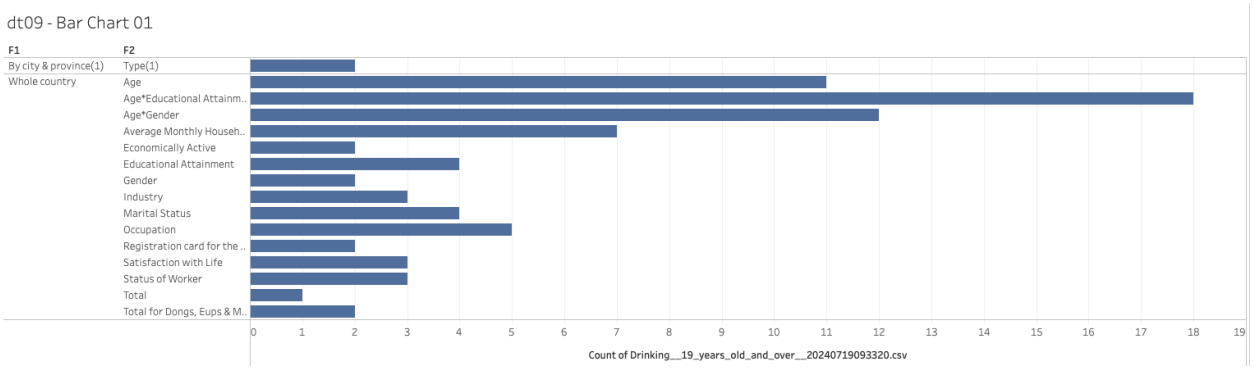
dt08 - Bubble Chart

03.41

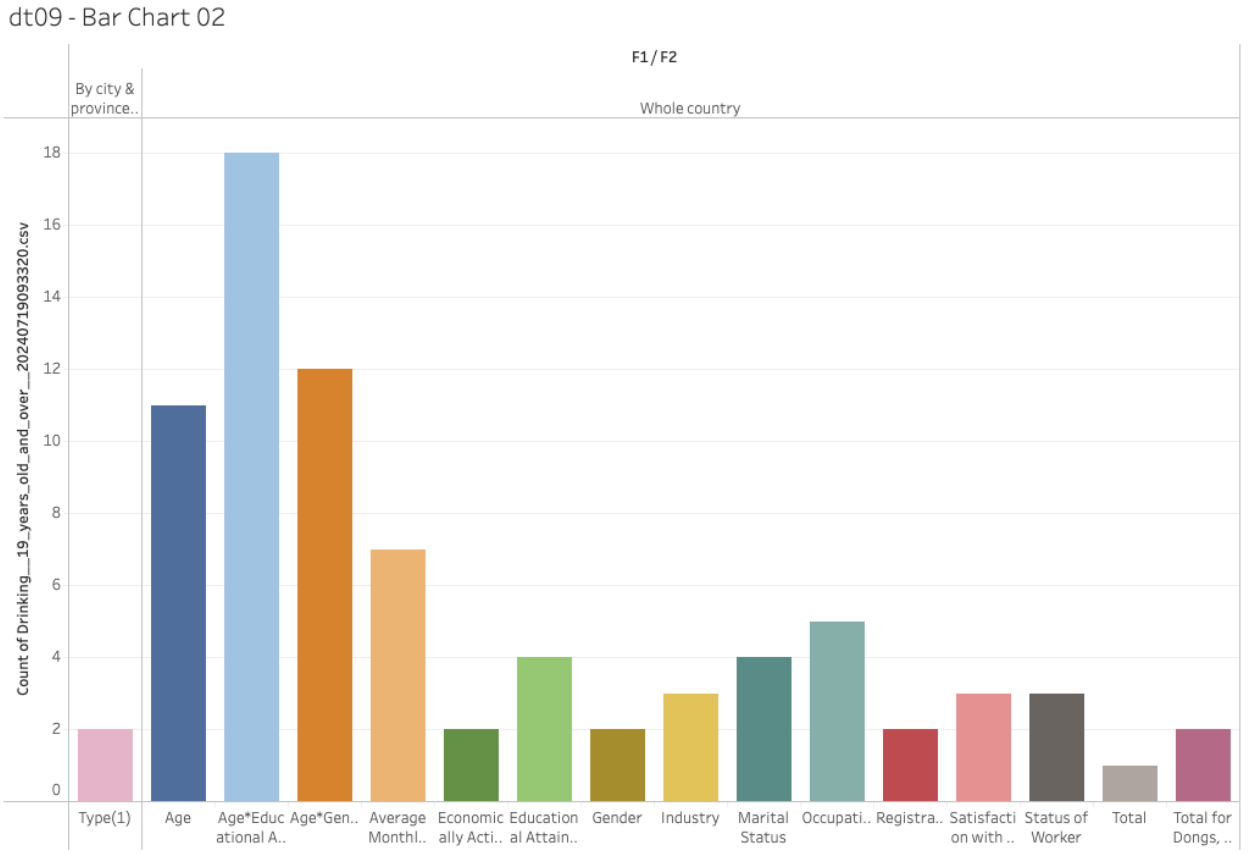
dt08 - Tree Map



03.42


dt09 - Bar Chart 01
F1
F2
By city & province(1)
Type(1)
Whole country
Age
Age*Educational Attainm..
Age*Gender
Average Monthly Househ..
Economically Active
Educational Attainment
Gender
Industry
Marital Status
Occupation
Registration card for the ..
Satisfaction with Life
Status of Worker
Total
Total for Dongs, Eups & M..
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Count of Drinking__19_years_old_and_over__20240719093320.csv

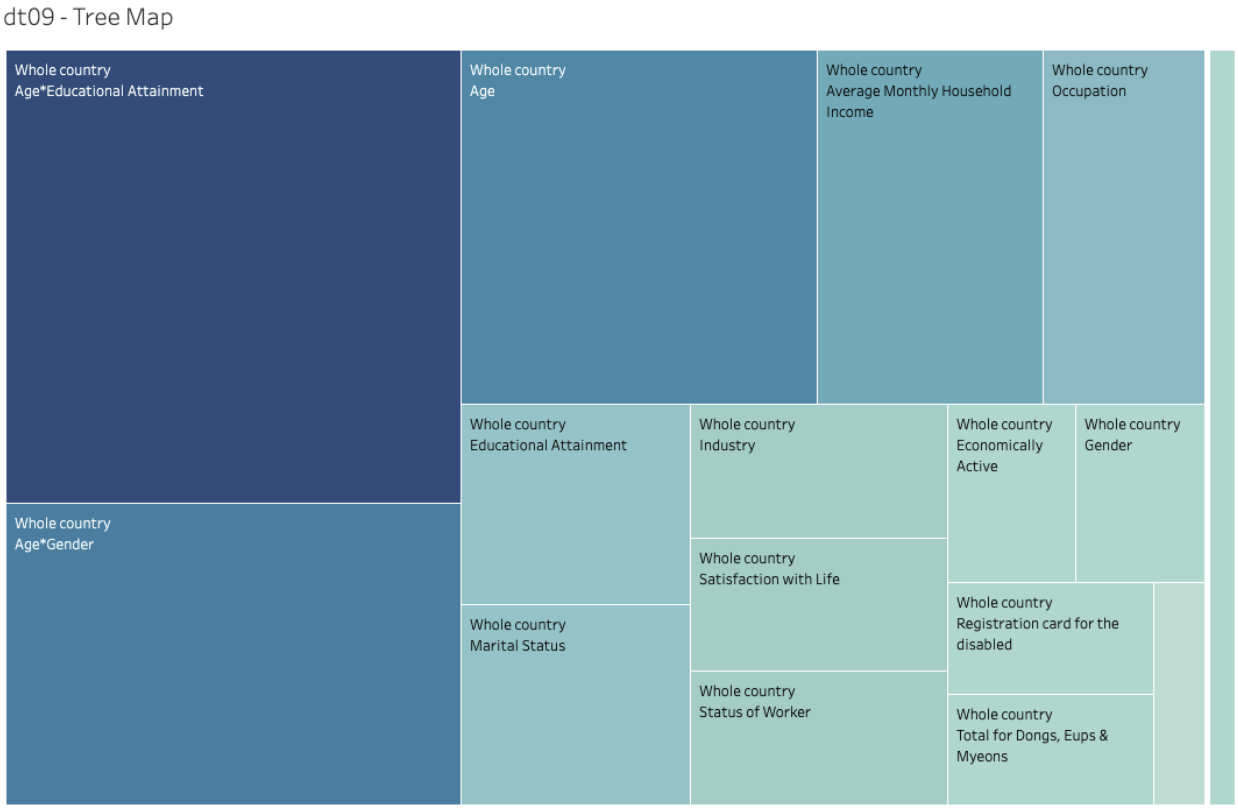
03.43


dt09 - Bar Chart 02
F1 / F2
By city & province..
Whole country
Count of Drinking__19_years_old_and_over__20240719093320.csv
0 2 4 6 8 10 12 14 16 18
Type(1)
Age
Age*Educational A..
Age*Gen..
Average Monthl..
Economically Acti..
Education al Attain..
Gender
Industry
Marital Status
Occupati..
Registra..
Satisfacti on with ..
Status of Worker
Total
Total for Dongs, ..

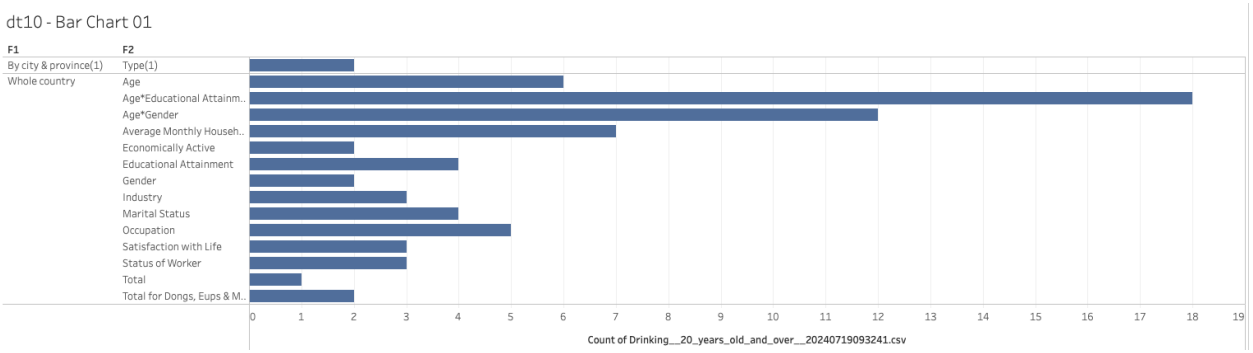
03.44


dt09 - Bubble Chart
Whole country Satisfaction
Whole country Status of
Whole country
Whole country
Whole country Average Monthly Household Income
Whole country
Whole country Educational Attainment
Whole country Age*Gender
By city & province(1)
Whole country Age
Whole country
Whole country Industry
Whole country Marital Status
Whole country Age*Educational Attainment
Whole country Occupation

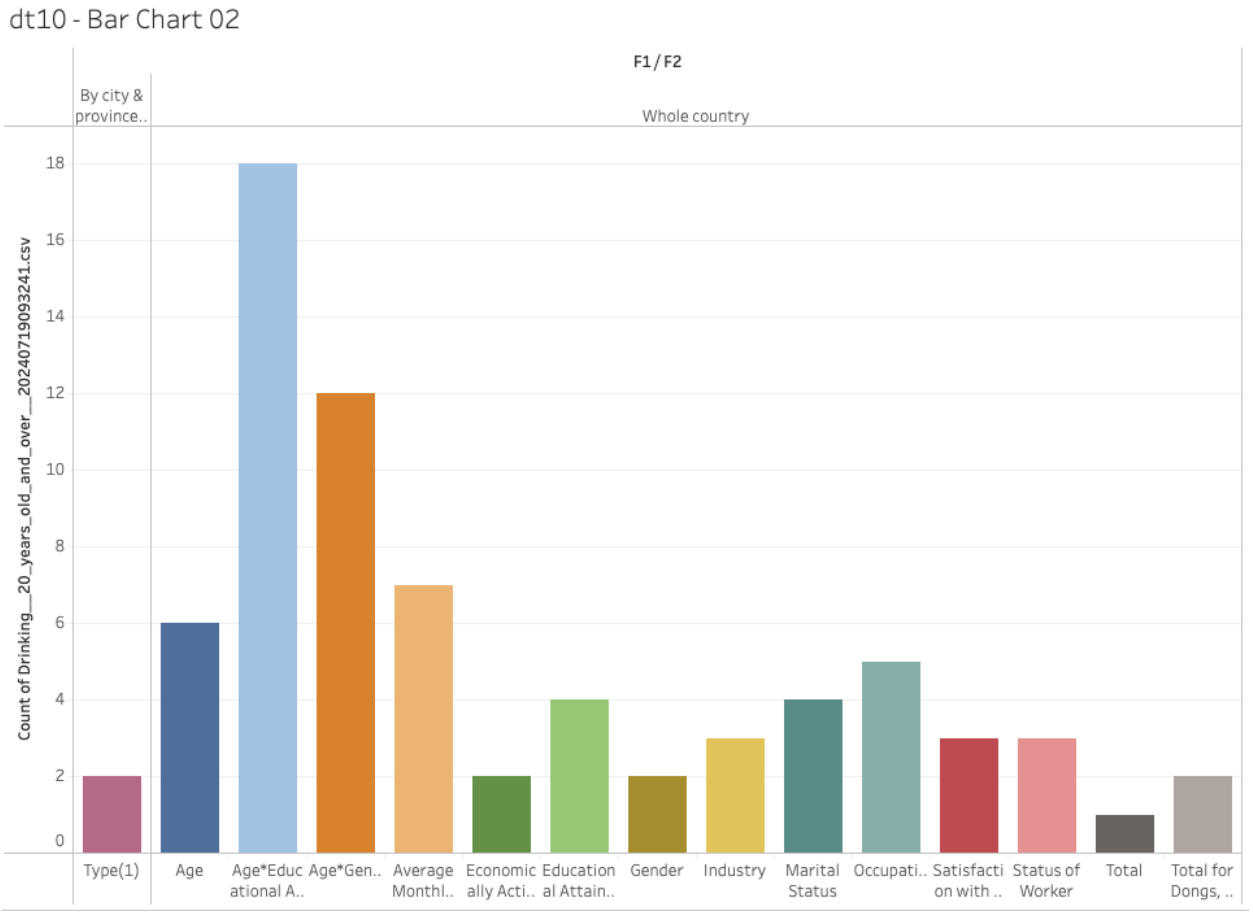
03.45


dt09 - Tree Map
Whole country Age*Educational Attainment
Whole country Age
Whole country Average Monthly Household Income
Whole country Occupation
Whole country Age*Gender
Whole country Educational Attainment
Whole country Industry
Whole country Economically Active
Whole country Gender
Whole country Satisfaction with Life
Whole country Marital Status
Whole country Registration card for the disabled
Whole country Status of Worker
Whole country Total for Dongs, Eups & Myeons

03.46


dt10 - Bar Chart 01
F1
F2
By city & province(1)
Type(1)
Whole country
Age
Age*Educational Attainm..
Age*Gender
Average Monthly Househ..
Economically Active
Educational Attainment
Gender
Industry
Marital Status
Occupation
Satisfaction with Life
Status of Worker
Total
Total for Dongs, Eups & M..
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Count of Drinking__20_years_old_and_over__20240719093241.csv

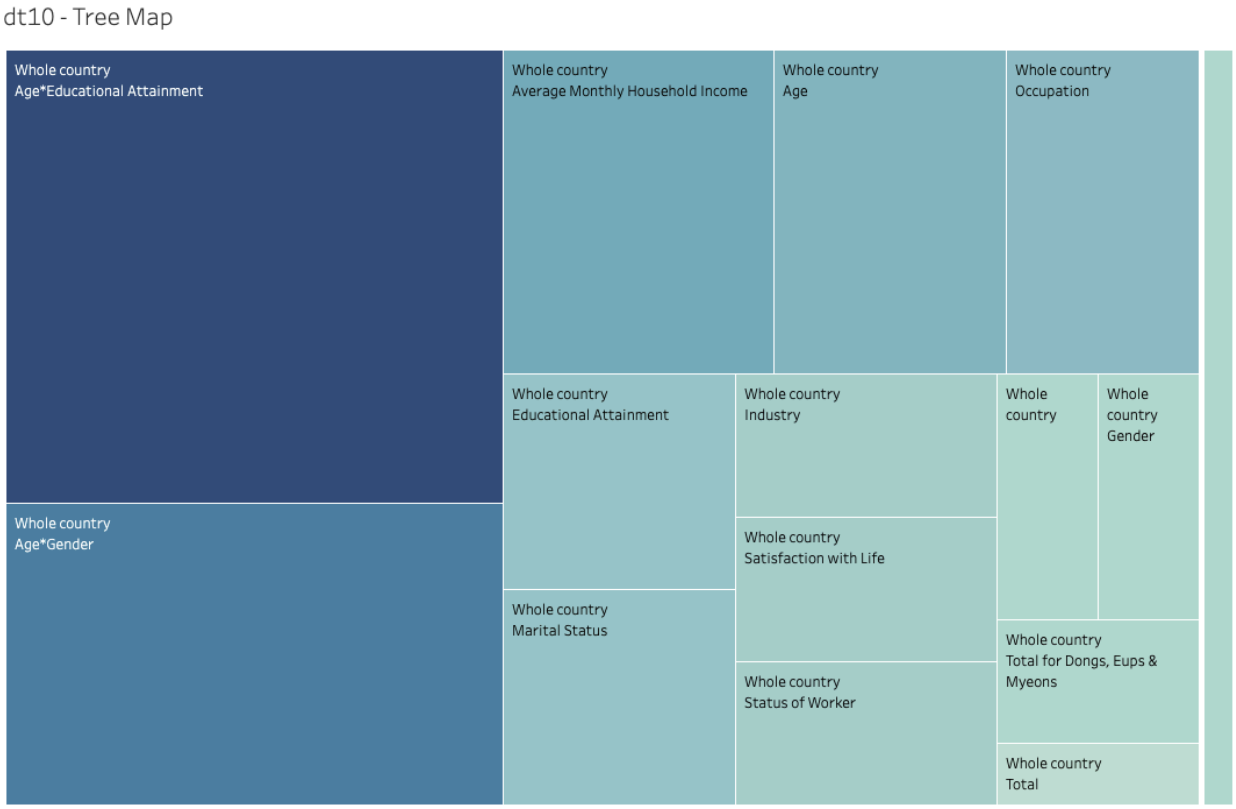
03.47


dt10 - Bar Chart 02
F1 / F2
By city & province..
Whole country
Count of Drinking__20_years_old_and_over__20240719093241.csv
0 2 4 6 8 10 12 14 16 18
Type(1)
Age
Age*Educational A..
Age*Gen..
Average Monthl..
Economically Acti..
Educational Attain..
Gender
Industry
Marital Status
Occupati..
Satisfaction with ..
Status of Worker
Total
Total for Dongs, ..

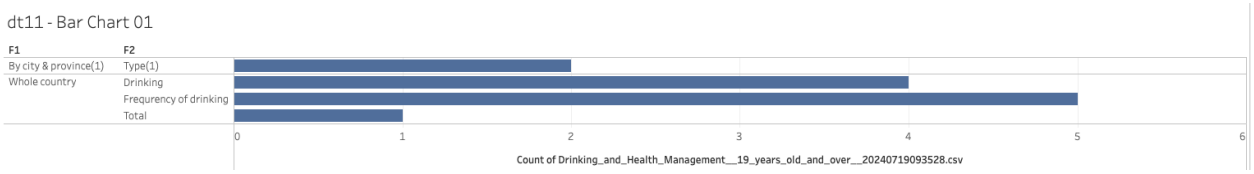
03.48


dt10 - Bubble Chart
Whole country Satisfaction with Life
Whole country Status of Worker
Whole country
Whole country Occupation
Whole country Average Monthly Household Income
Whole country
Whole country Educational Attainment
Whole country Age*Gender
By city & province(1)
Whole country Age
Whole country
Whole country Industry
Whole country Age*Educational Attainment
Whole country Marital Status

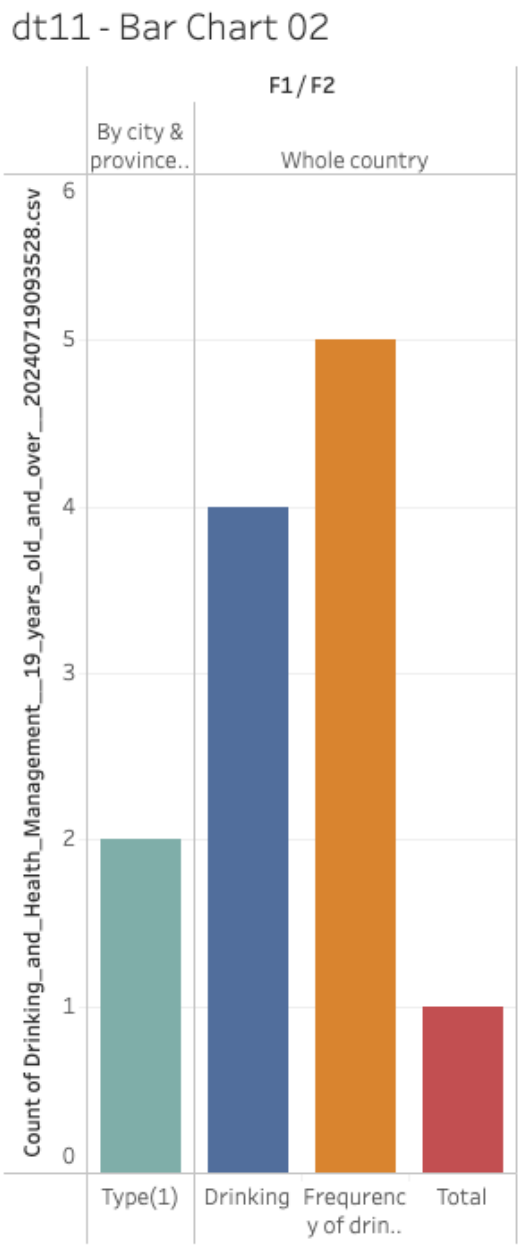
03.49


dt10 - Tree Map
Whole country Age*Educational Attainment
Whole country Average Monthly Household Income
Whole country Age
Whole country Occupation
Whole country Educational Attainment
Whole country Industry
Whole country
Whole country Gender
Whole country Age*Gender
Whole country Satisfaction with Life
Whole country Marital Status
Whole country Total for Dongs, Eups & Myeons
Whole country Status of Worker
Whole country Total

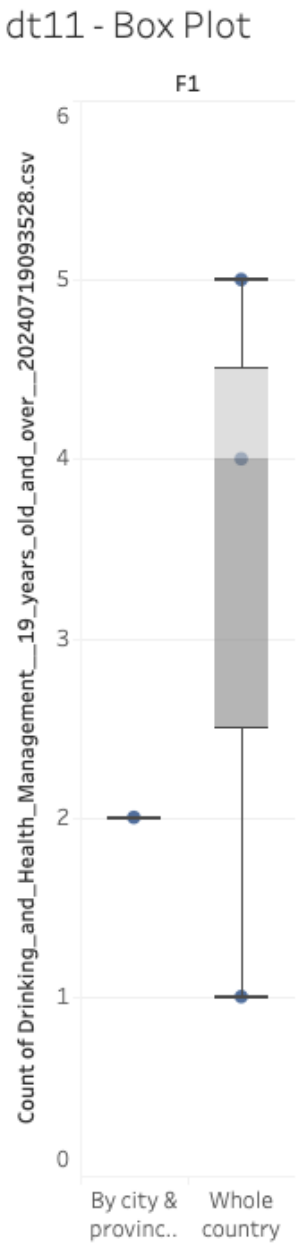
03.50


dt11 - Bar Chart 01
F1
F2
By city & province(1)
Type(1)
Whole country
Drinking
Frequency of drinking
Total
0
1
2
3
4
5
6
Count of Drinking_and_Health_Management__19_years_old_and_over__20240719093528.csv

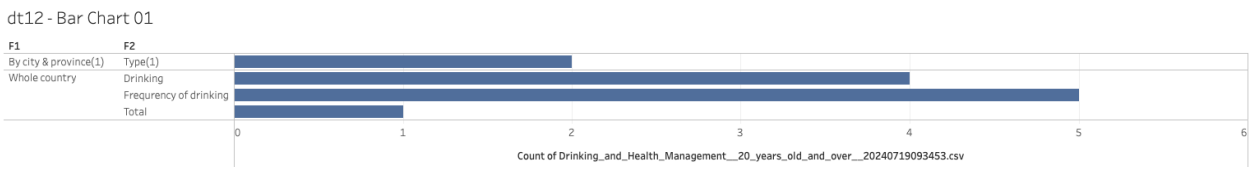
03.51


dt11 - Bar Chart 02
F1 / F2
By city & province..
Whole country
Count of Drinking_and_Health_Management__19_years_old_and_over__20240719093528.csv
6
5
4
3
2
1
0
Type(1)
Drinking
Frequenc y of drin..
Total

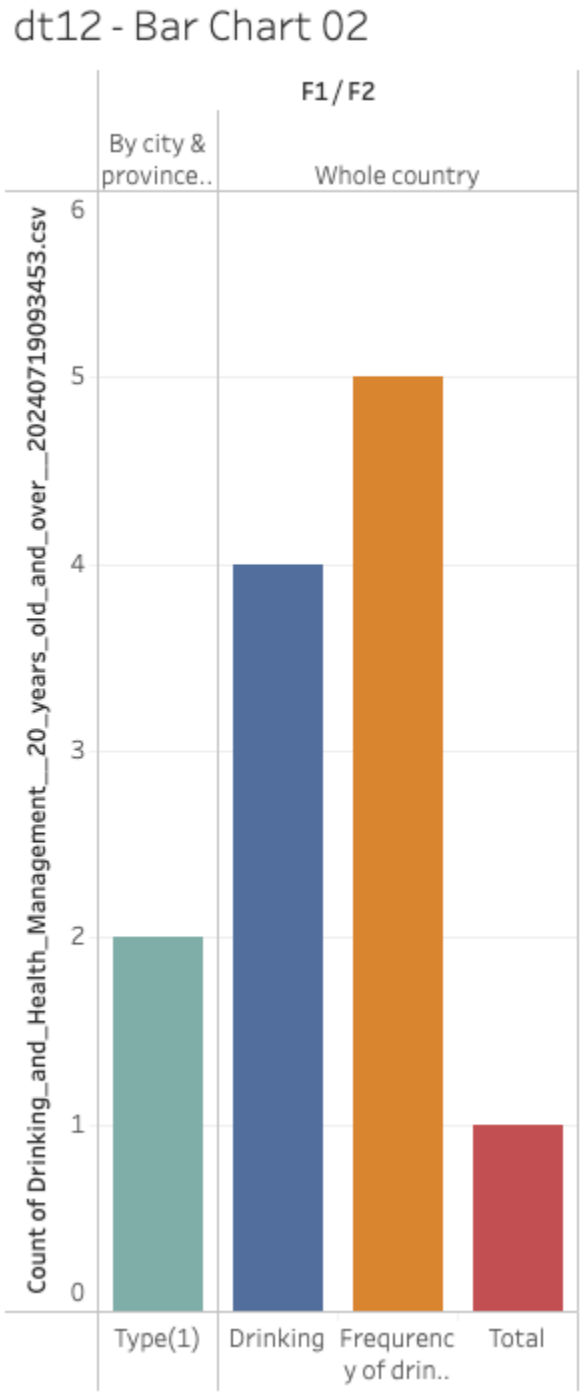
03.52


dt11 - Box Plot
F1
Count of Drinking_and_Health_Management__19_years_old_and_over__20240719093528.csv
6
5
4
3
2
1
0
By city & provinc..
Whole country

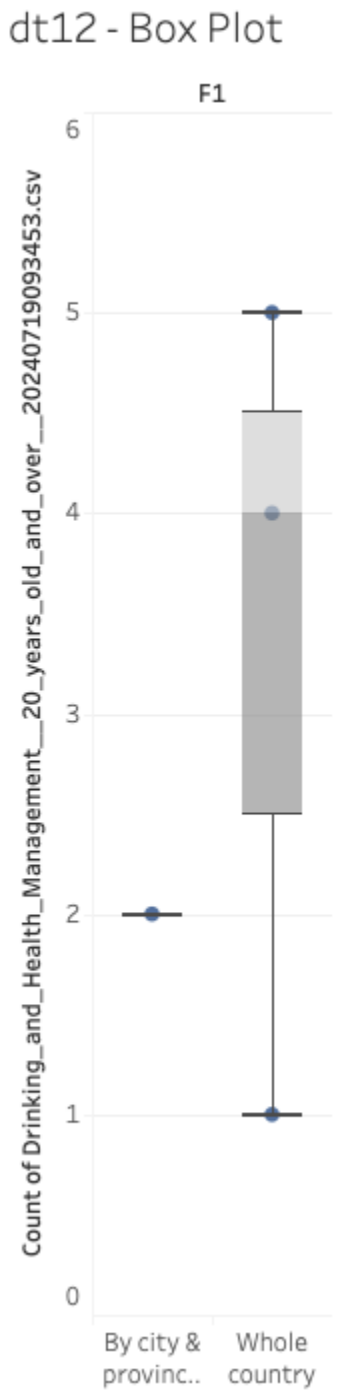
03.53


dt12 - Bar Chart 01
F1
F2
By city & province(1)
Type(1)
Whole country
Drinking
Frequency of drinking
Total
0
1
2
3
4
5
6
Count of Drinking_and_Health_Management__20_years_old_and_over__20240719093453.csv

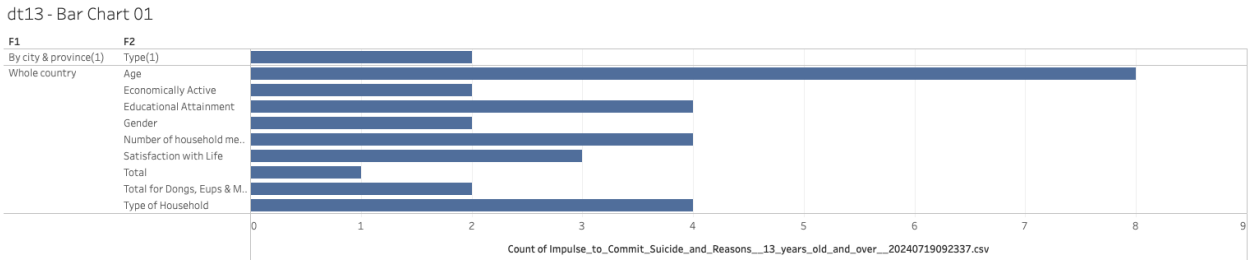
03.54


dt12 - Bar Chart 02
F1 / F2
By city & province..
Whole country
Count of Drinking_and_Health_Management__20_years_old_and_over__20240719093453.csv
6
5
4
3
2
1
0
Type(1)
Drinking
Frequency of drin..
Total

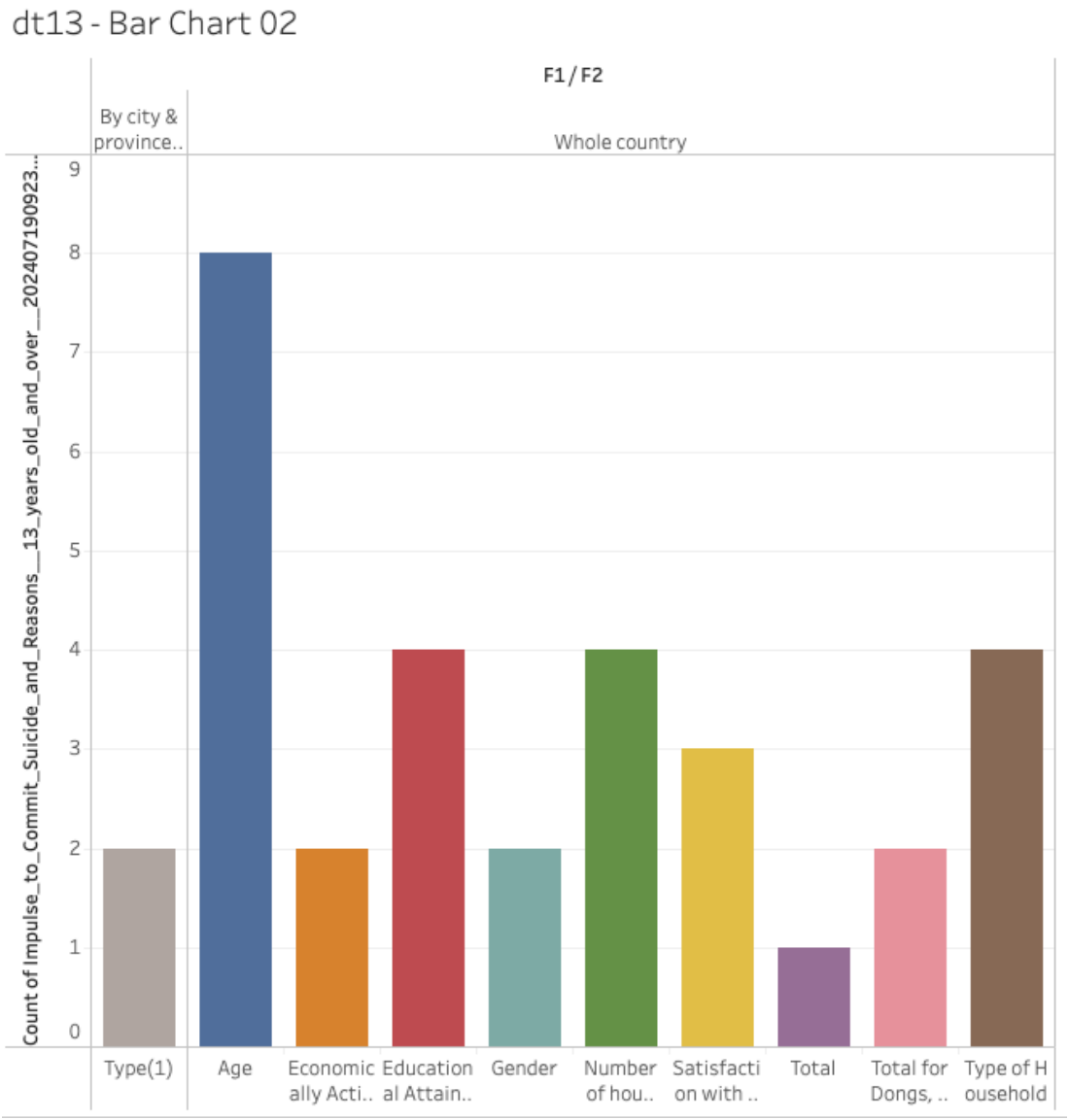
03.55


dt12 - Box Plot
F1
Count of Drinking_and_Health_Management__20_years_old_and_over__20240719093453.csv
6
5
4
3
2
1
0
By city & provinc..
Whole country

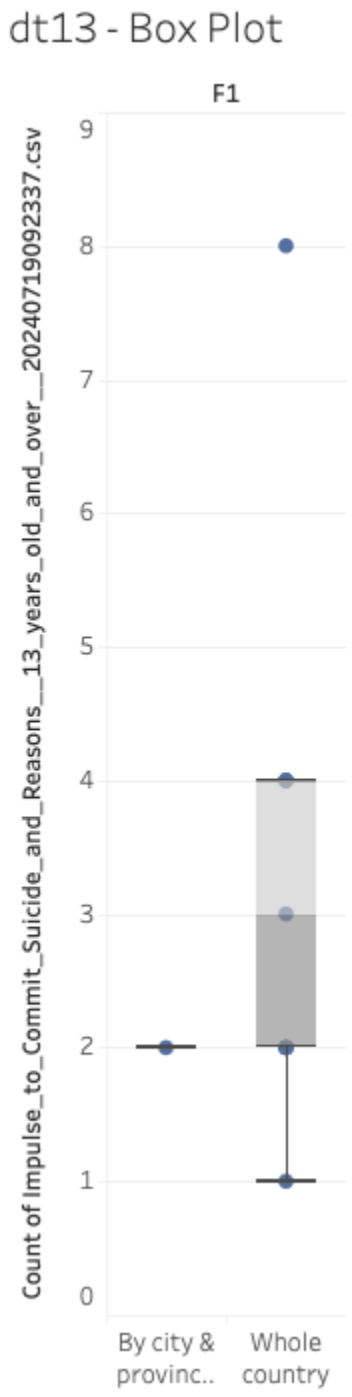
03.56


dt13 - Bar Chart 01
F1
F2
By city & province(1)
Whole country
Type(1)
Age
Economically Active
Educational Attainment
Gender
Number of household me..
Satisfaction with Life
Total
Total for Dongs, Eups & M..
Type of Household
0
1
2
3
4
5
6
7
8
9
Count of Impulse_to_Commit_Suicide_and_Reasons__13_years_old_and_over__20240719092337.csv

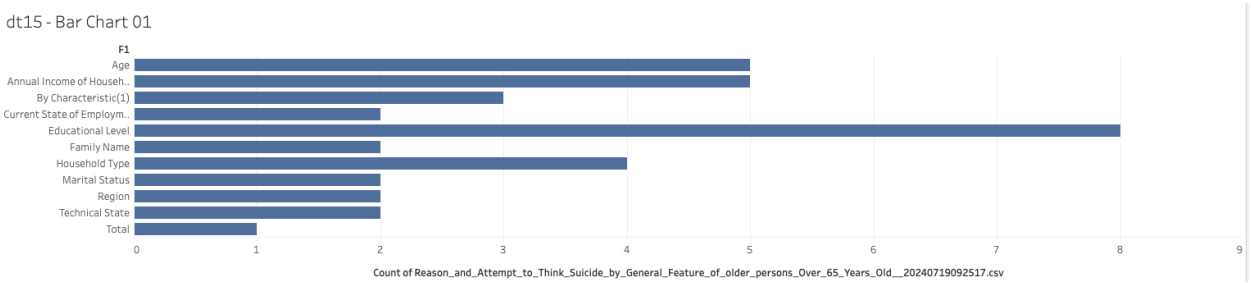
03.57


dt13 - Bar Chart 02
F1 / F2
By city & province..
Whole country
Count of Impulse_to_Commit_Suicide_and_Reasons__13_years_old_and_over__20240719092...
0
1
2
3
4
5
6
7
8
9
Type(1)
Age
Economically Acti..
Educational Attain..
Gender
Number of hou..
Satisfaction with ..
Total
Total for Dongs, ..
Type of Household

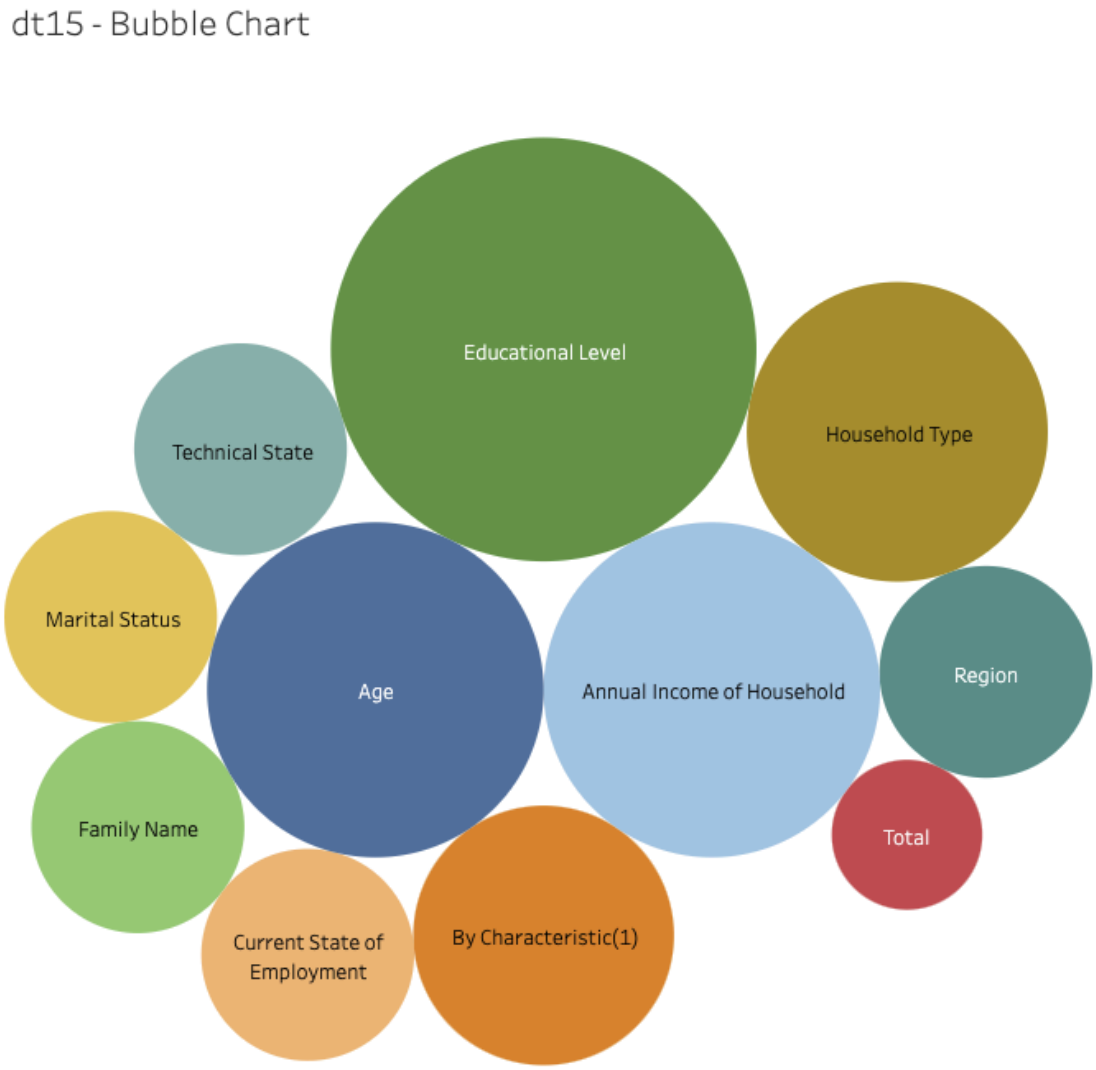
03.58


dt13 - Box Plot
F1
Count of Impulse_to_Commit_Suicide_and_Reasons__13_years_old_and_over__20240719092337.csv
9
8
7
6
5
4
3
2
1
0
By city & provinc..
Whole country

03.59

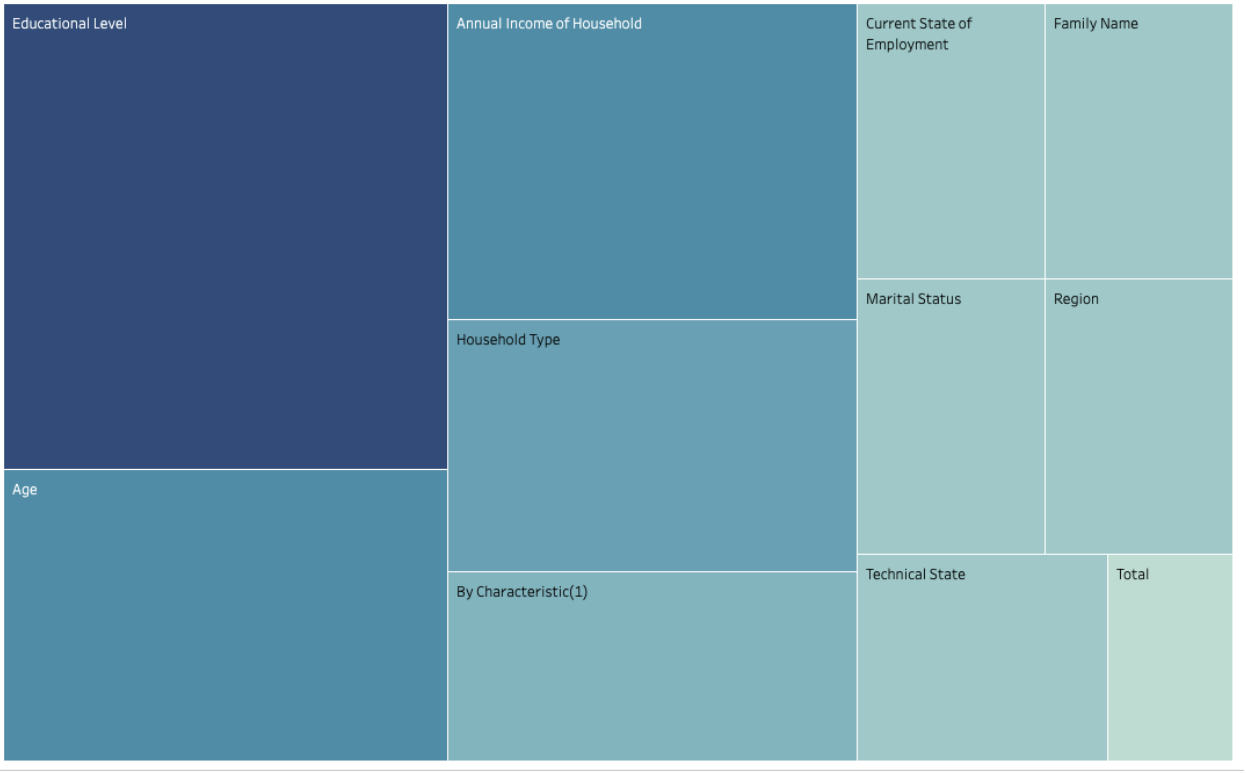

dt15 - Bar Chart 01
F1
Age
Annual Income of Househ..
By Characteristic(1)
Current State of Employm..
Educational Level
Family Name
Household Type
Marital Status
Region
Technical State
Total
0
1
2
3
4
5
6
7
8
9
Count of Reason_and_Attempt_to_Think_Suicide_by_General_Feature_of_older_persons_Over_65_Years_Old__20240719092517.csv

03.60


dt15 - Bubble Chart
Educational Level
Technical State
Household Type
Marital Status
Age
Annual Income of Household
Region
Family Name
Total
Current State of Employment
By Characteristic(1)

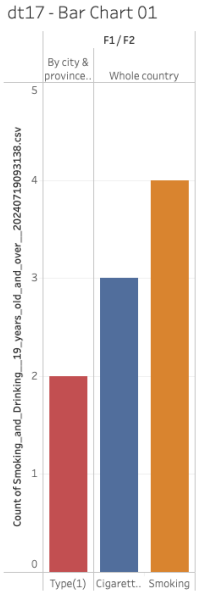
03.61

dt15 - Tree Map

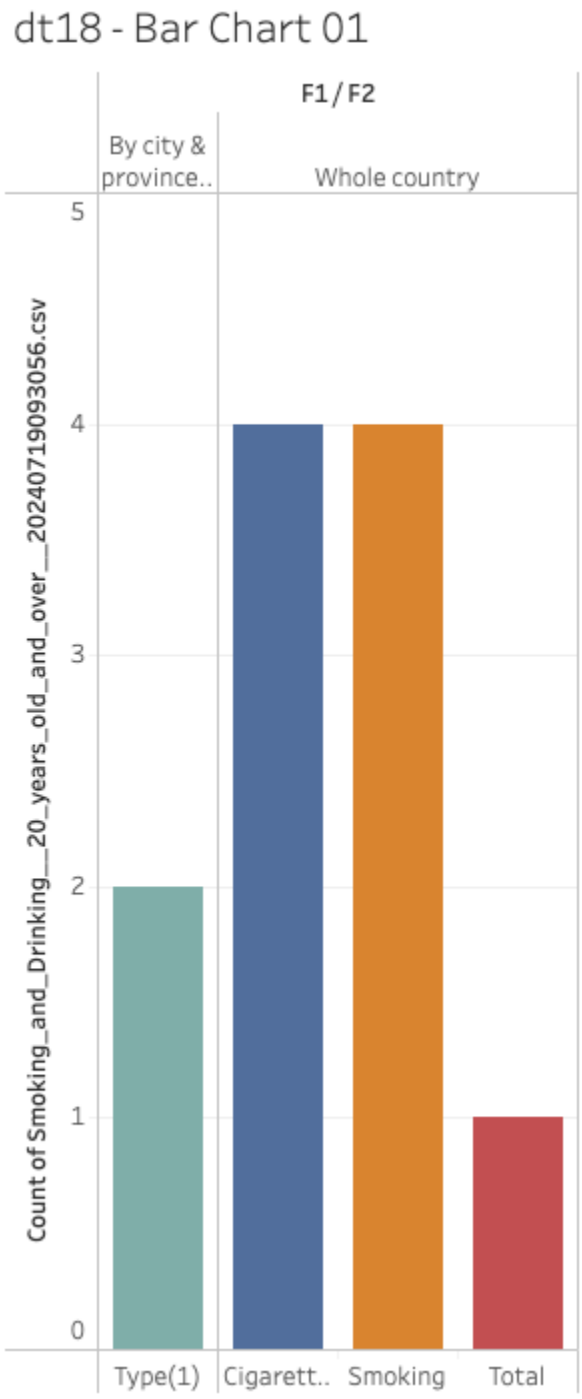


03.62

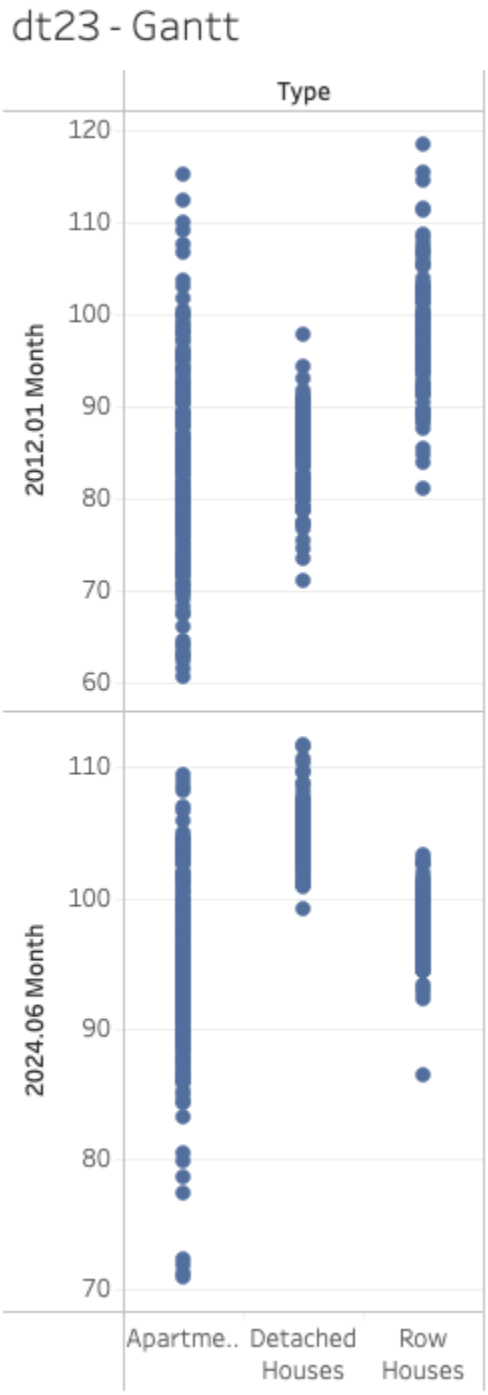
dt17 - Bar Chart 01



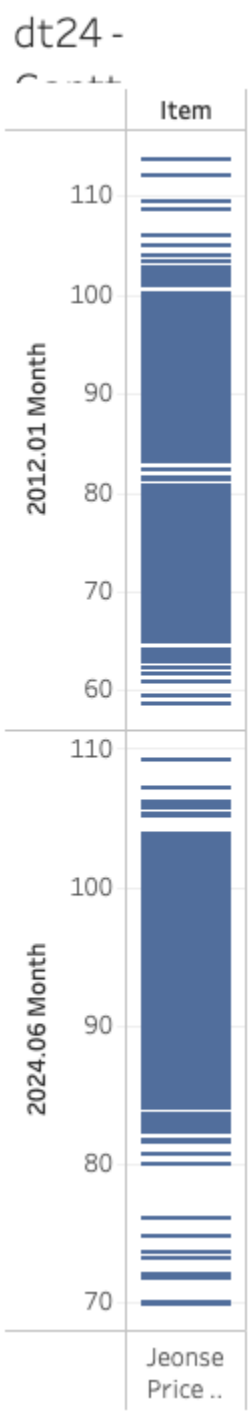
03.63


dt18 - Bar Chart 01
F1 / F2
By city & province..
Whole country
Count of Smoking_and_Drinking__20_years_old_and_over__20240719093056.csv
5
4
3
2
1
0
Type(1)
Cigarett..
Smoking
Total

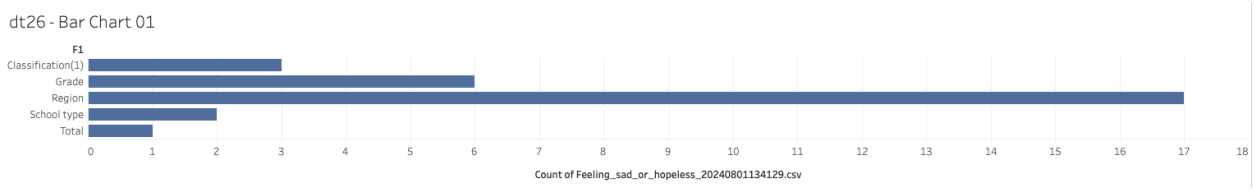
03.64


dt23 - Gantt
Type
2012.01 Month
120
110
100
90
80
70
60
2024.06 Month
110
100
90
80
70
Apartme..
Detached Houses
Row Houses

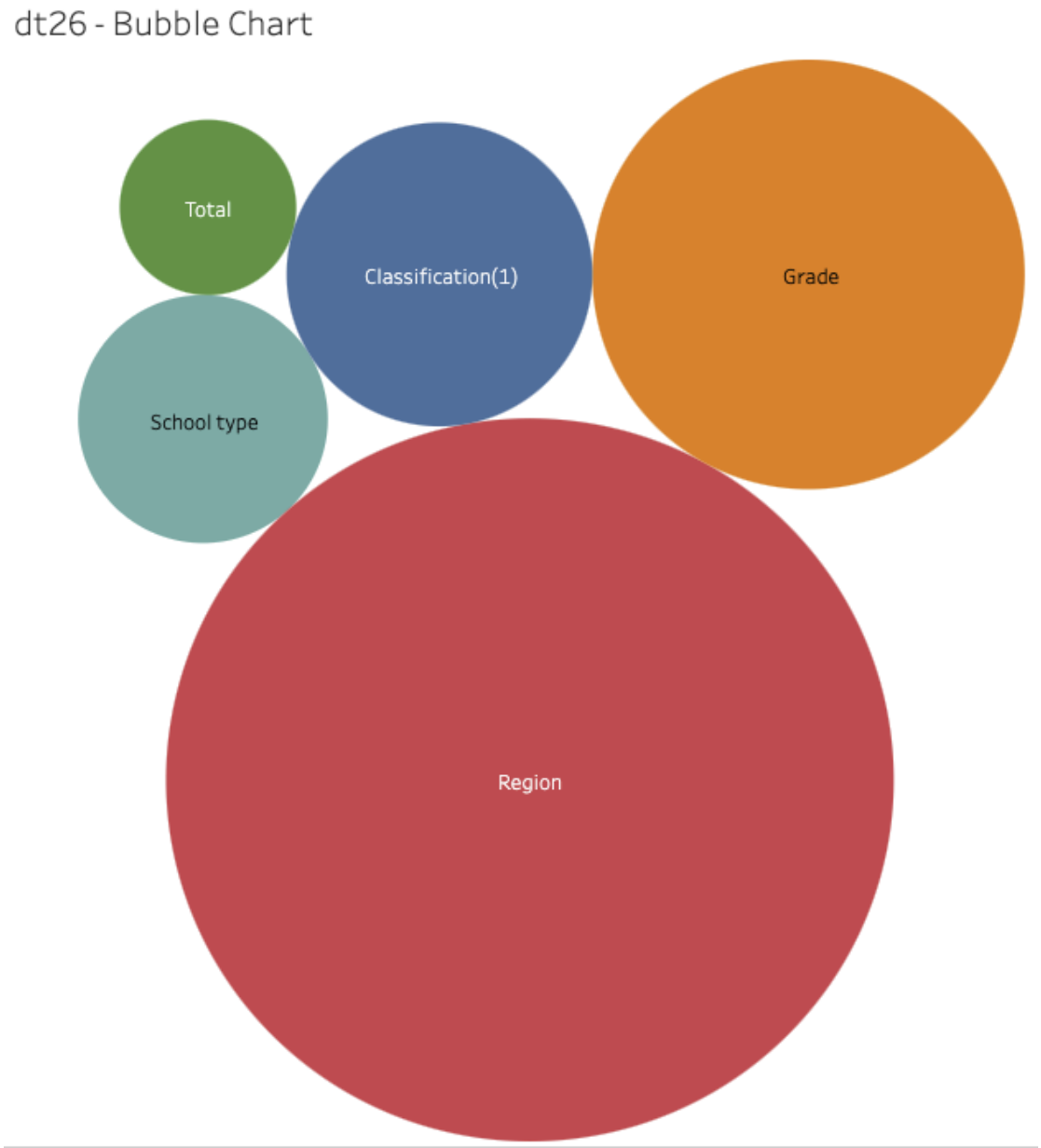
03.65


dt24 -
Item
2012.01 Month
110
100
90
80
70
60
2024.06 Month
110
100
90
80
70
Jeonse Price ..

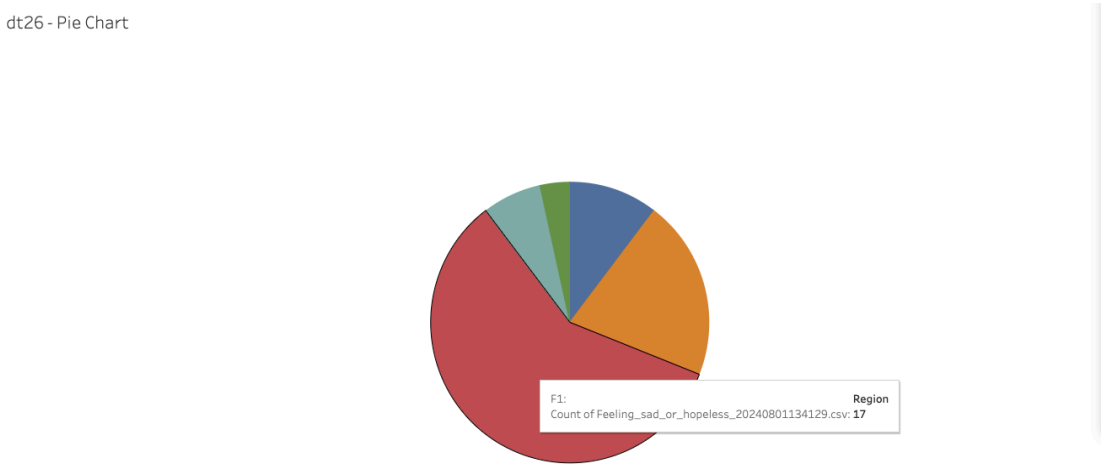
03.66


dt26 - Bar Chart 01
F1
Classification(1)
Grade
Region
School type
Total
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Count of Feeling_sad_or_hopeless_20240801134129.csv

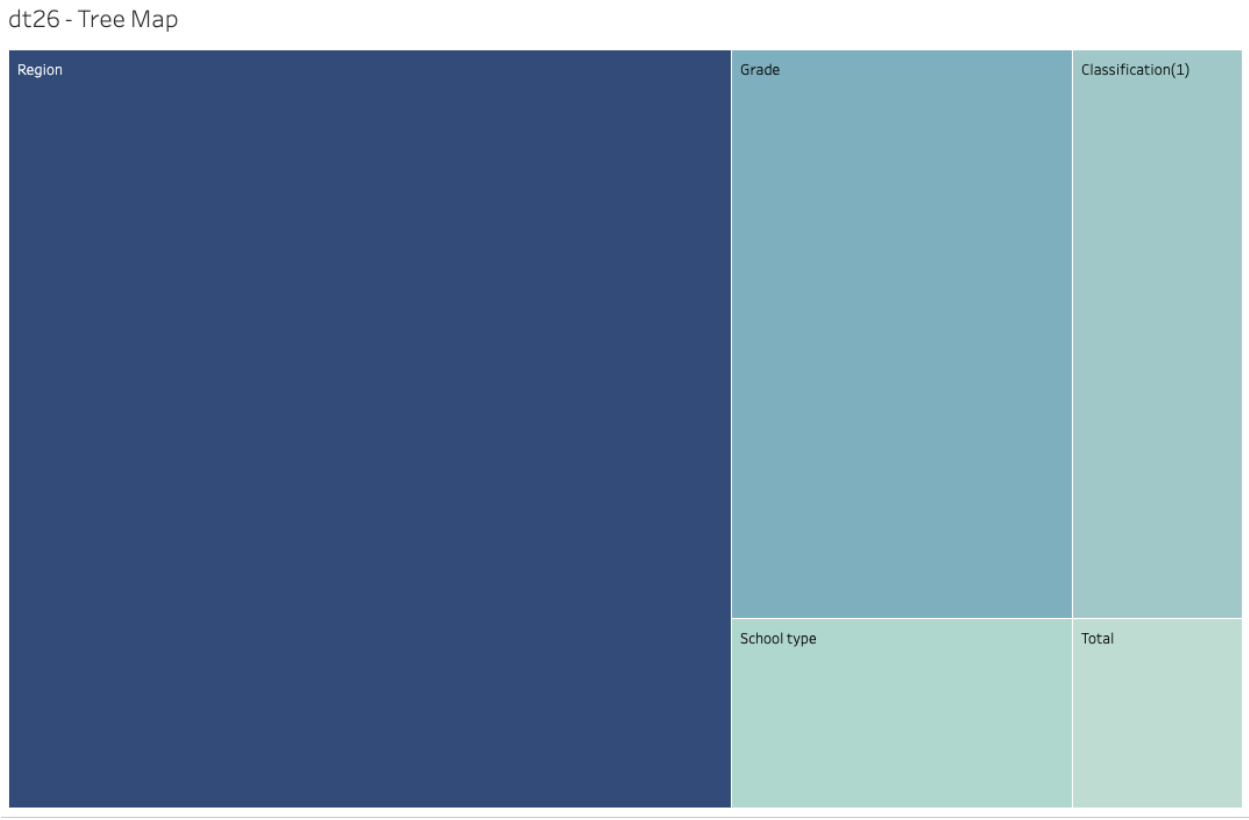
03.67


dt26 - Bubble Chart
Total
Classification(1)
Grade
School type
Region

03.68

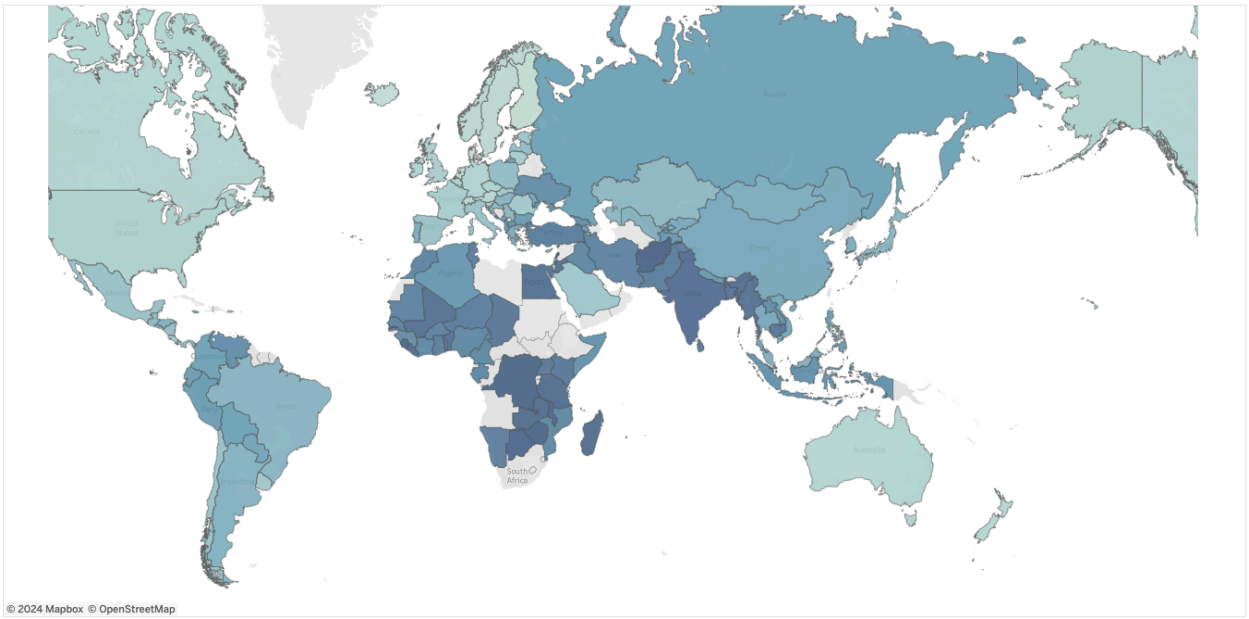

dt26 - Pie Chart
F1:
Region
Count of Feeling_sad_or_hopeless_20240801134129.csv: 17

03.69


dt26 - Tree Map
Region
Grade
Classification(1)
School type
Total

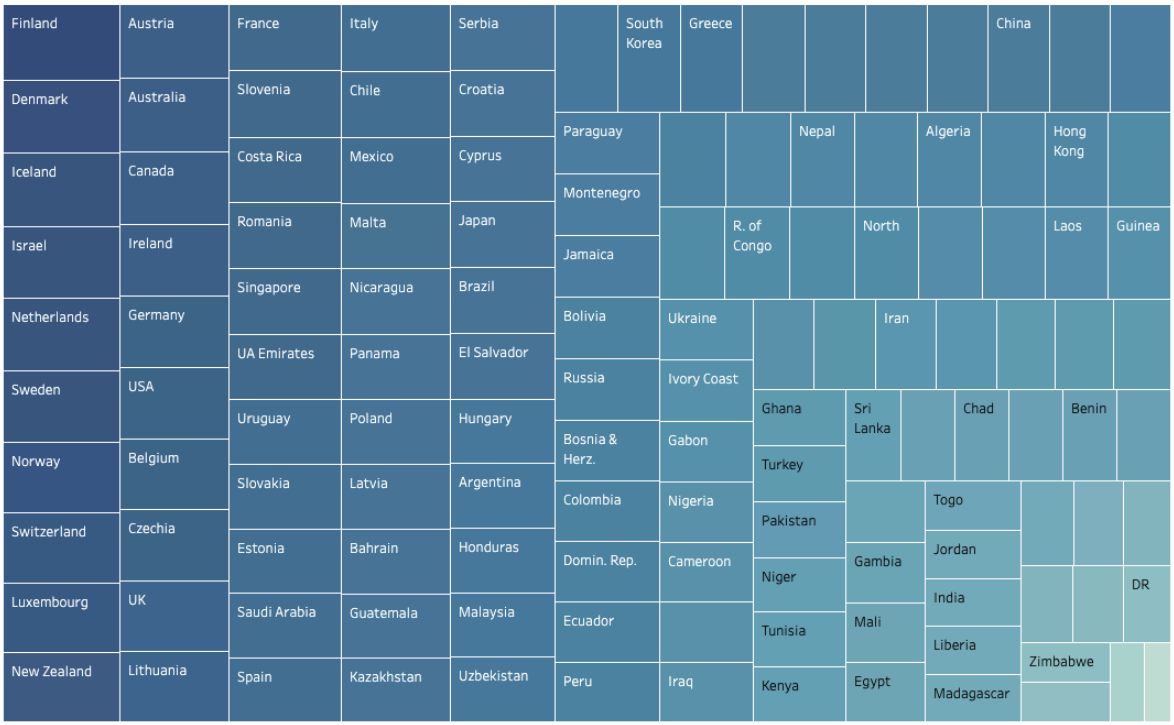
03.70

dt27 - Area Map

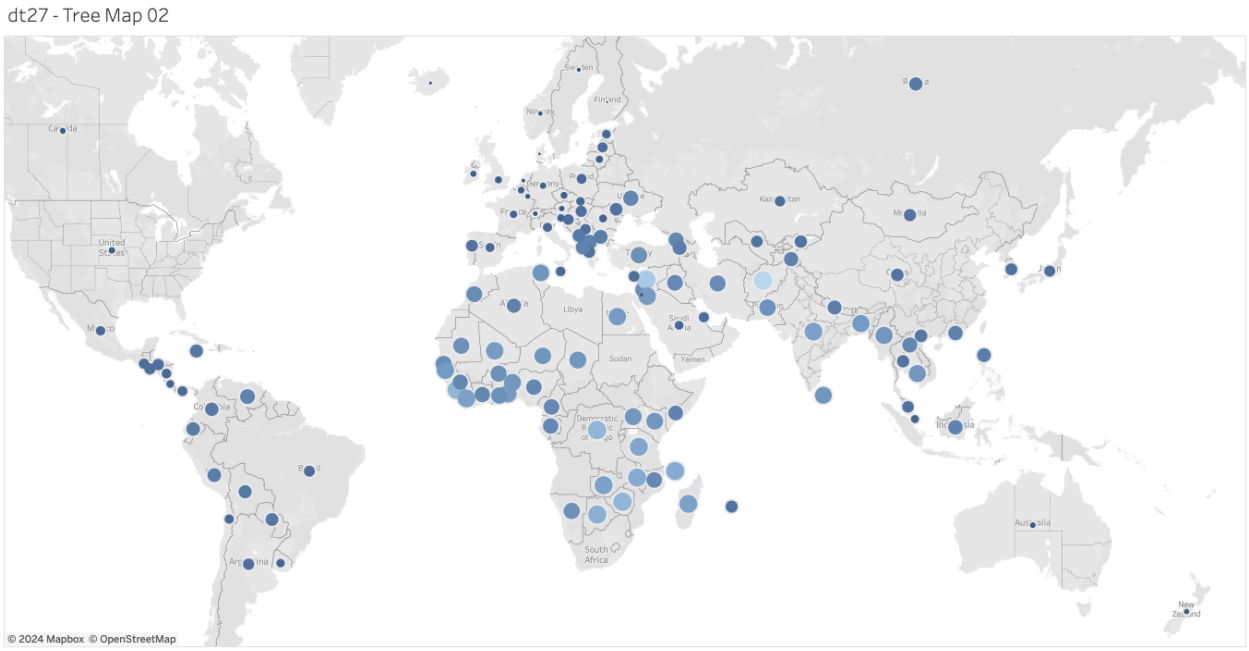


03.71

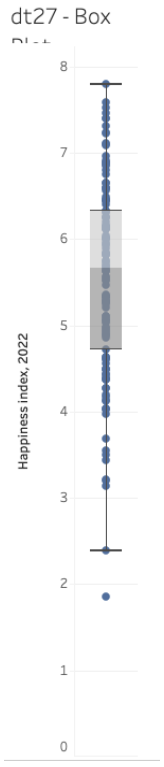
dt27 - Tree Map



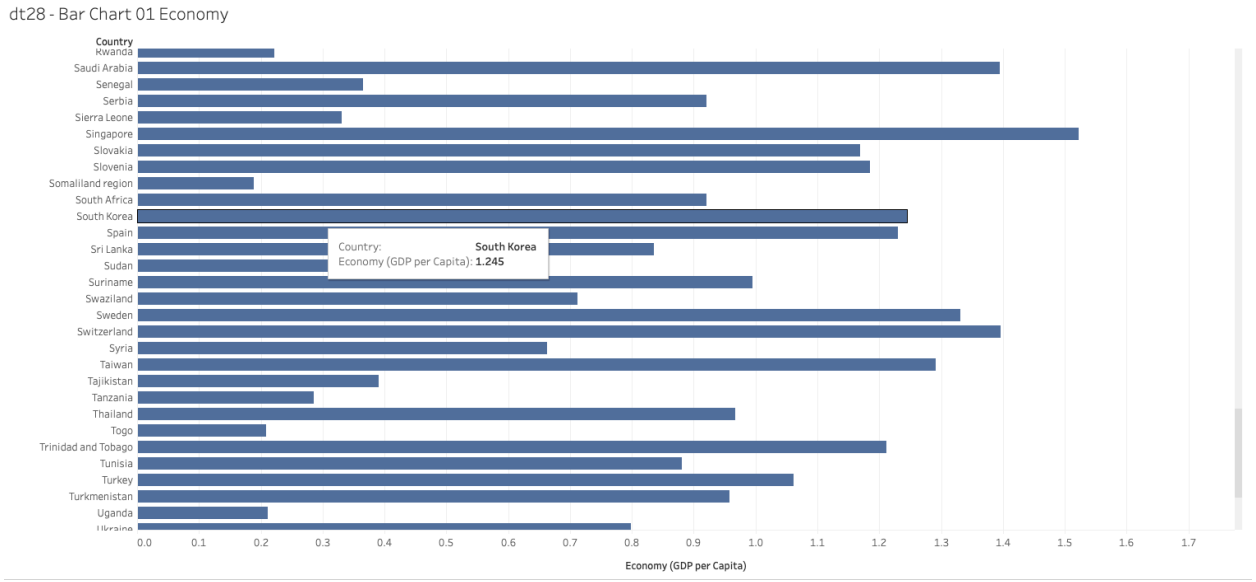
03.72


dt27 - Tree Map 02
© 2024 Mapbox © OpenStreetMap

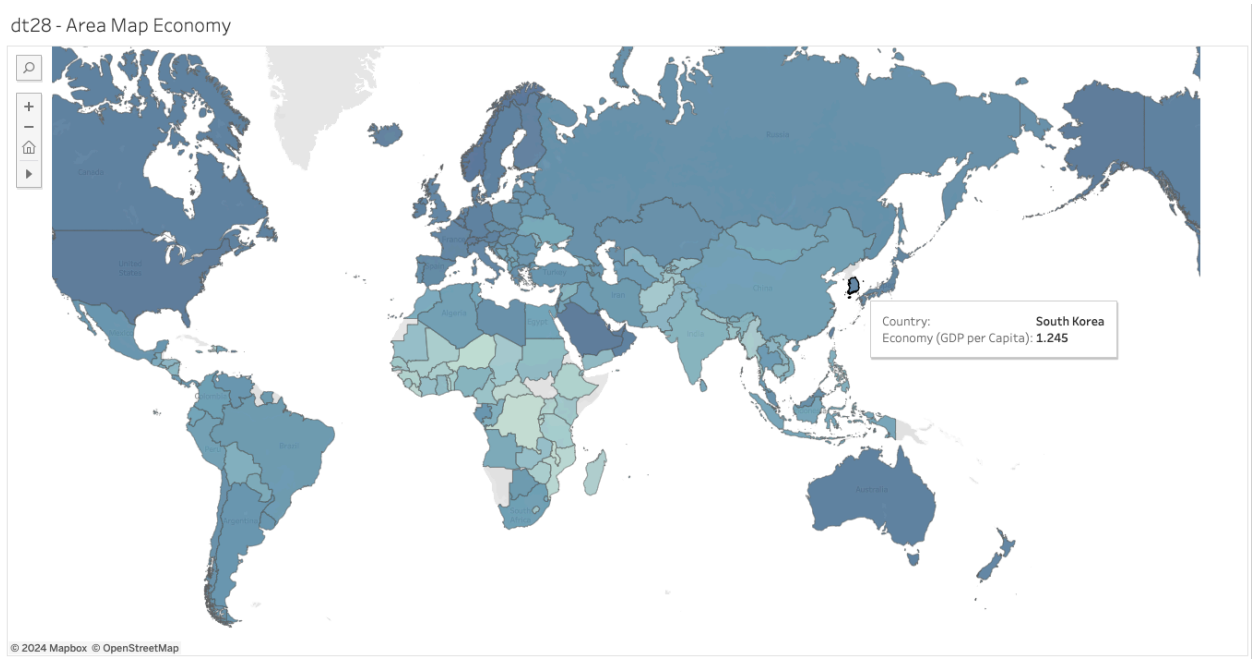
03.73


dt27 - Box
Happiness index, 2022
8
7
6
5
4
3
2
1
0

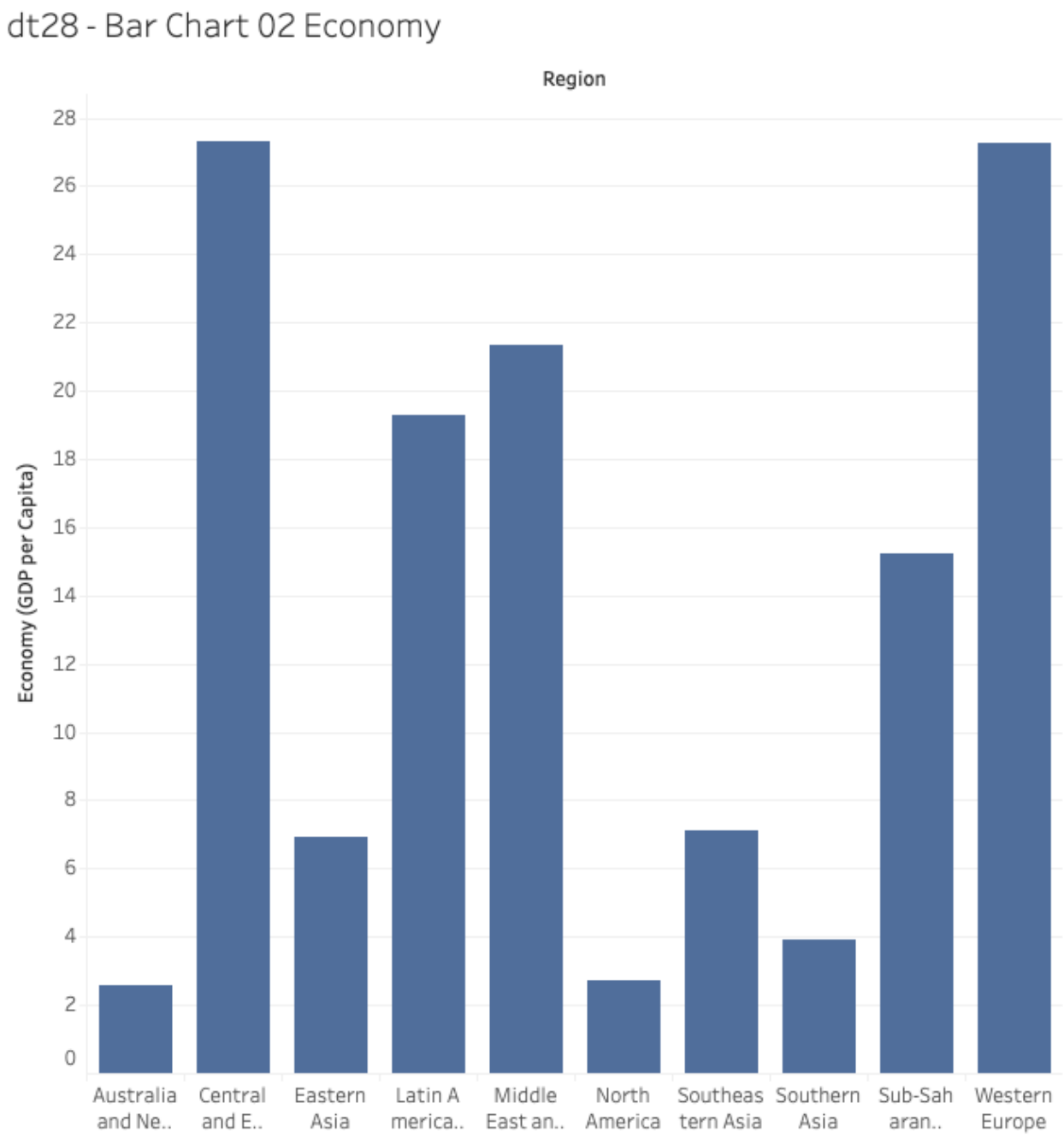
03.74


dt28 - Bar Chart 01 Economy
Country
Rwanda
Saudi Arabia
Senegal
Serbia
Sierra Leone
Singapore
Slovakia
Slovenia
Somaliland region
South Africa
South Korea
Spain
Sri Lanka
Sudan
Suriname
Swaziland
Sweden
Switzerland
Syria
Taiwan
Tajikistan
Tanzania
Thailand
Togo
Trinidad and Tobago
Tunisia
Turkey
Turkmenistan
Uganda
Ukraine
Country: South Korea
Economy (GDP per Capita): 1.245
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7
Economy (GDP per Capita)

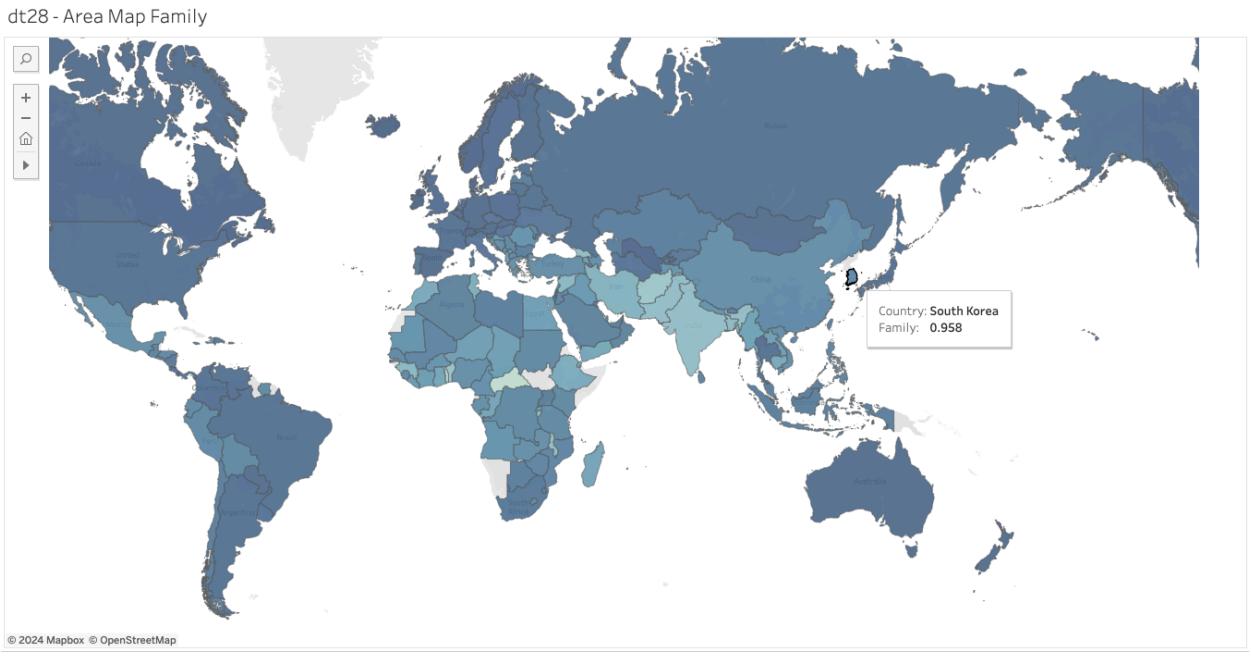
03.75


dt28 - Area Map Economy
Country: South Korea
Economy (GDP per Capita): 1.245
© 2024 Mapbox © OpenStreetMap

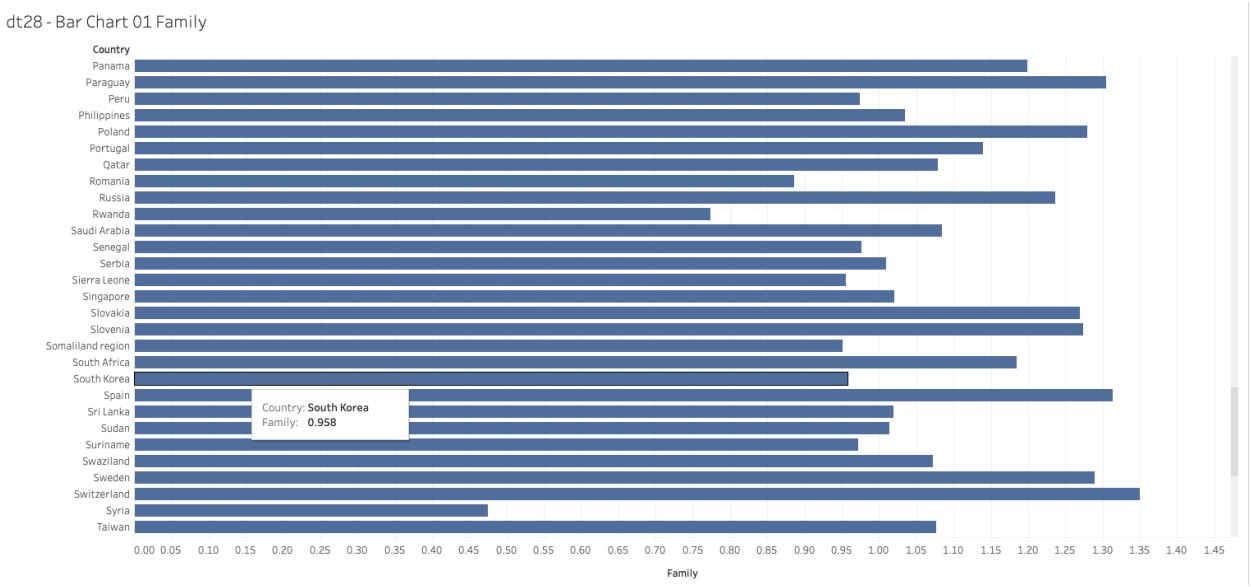
03.76


dt28 - Bar Chart 02 Economy
Region
Economy (GDP per Capita)
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
Australia and Ne..
Central and E..
Eastern Asia
Latin America..
Middle East an..
North America
Southeastern Asia
Southern Asia
Sub-Saharan..
Western Europe

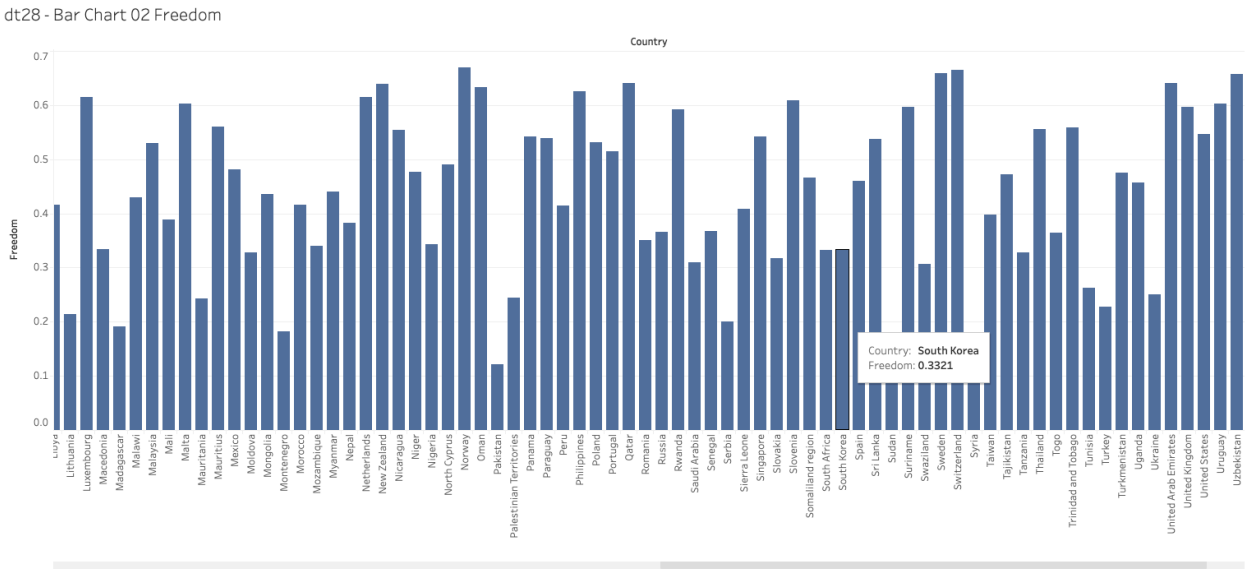
03.77


dt28 - Area Map Family
Country: South Korea
Family: 0.958
© 2024 Mapbox © OpenStreetMap

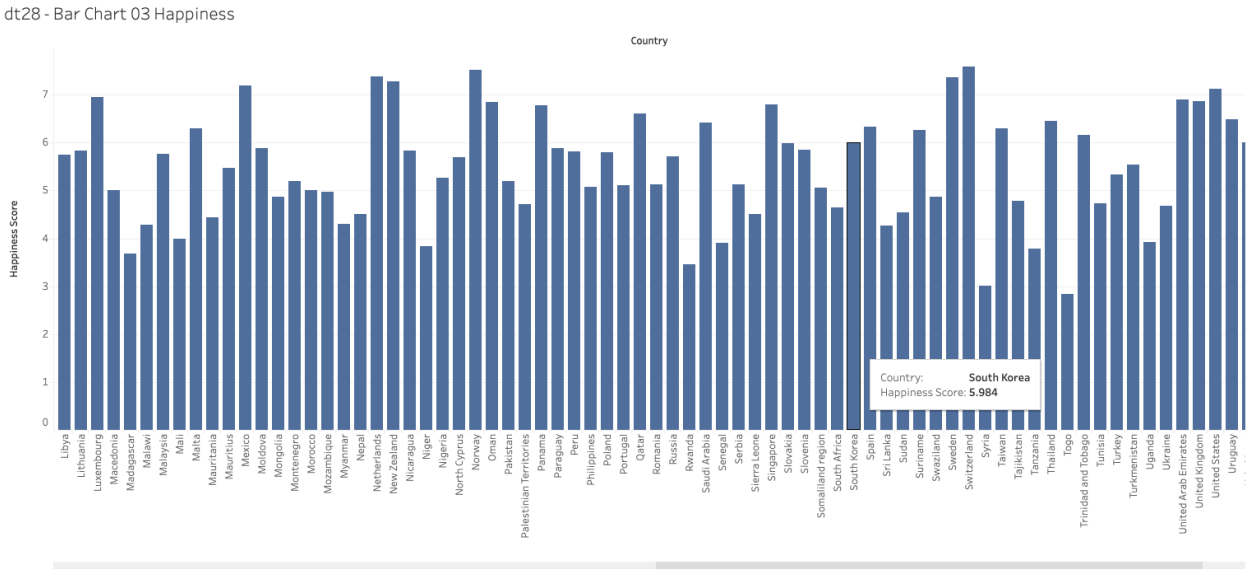
03.78


dt28 - Bar Chart 01 Family
Country
Panama
Paraguay
Peru
Philippines
Poland
Portugal
Qatar
Romania
Russia
Rwanda
Saudi Arabia
Senegal
Serbia
Sierra Leone
Singapore
Slovakia
Slovenia
Somaliland region
South Africa
South Korea
Spain
Sri Lanka
Sudan
Suriname
Swaziland
Sweden
Switzerland
Syria
Taiwan
Country: South Korea
Family: 0.958
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00 1.05 1.10 1.15 1.20 1.25 1.30 1.35 1.40 1.45
Family

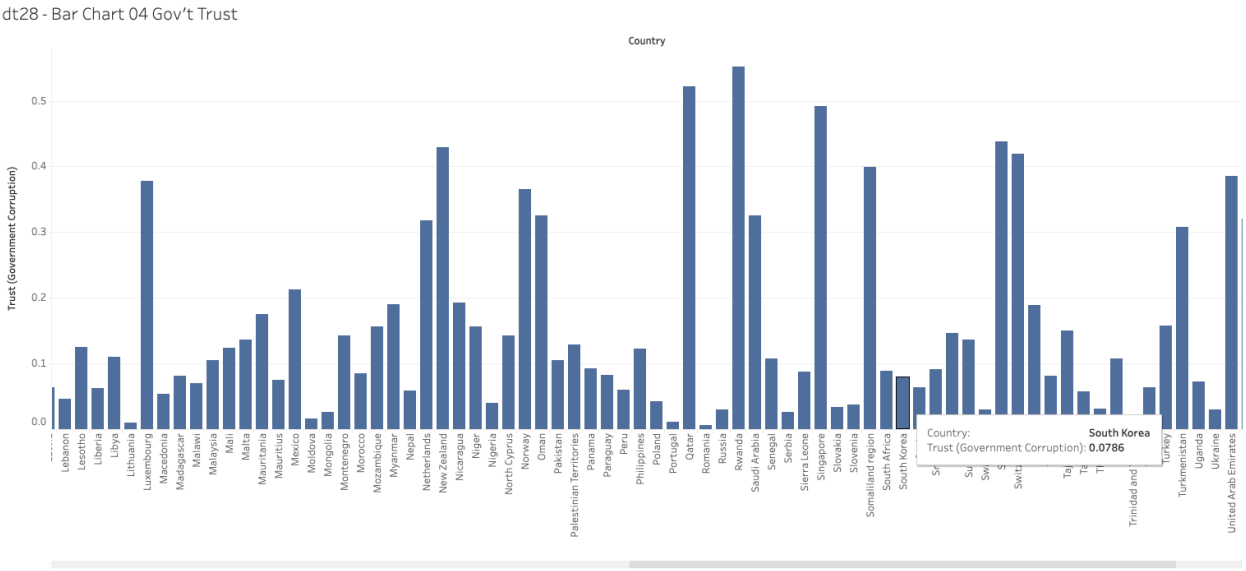
03.79


dt28 - Bar Chart 02 Freedom
Country
Freedom
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Country: South Korea
Freedom: 0.3321
Lithuania
Luxembourg
Macedonia
Madagascar
Malawi
Malaysia
Mali
Malta
Mauritania
Mauritius
Mexico
Moldova
Mongolia
Montenegro
Morocco
Mozambique
Myanmar
Nepal
Netherlands
New Zealand
Nicaragua
Niger
Nigeria
North Cyprus
Norway
Oman
Pakistan
Palestinian Territories
Panama
Paraguay
Peru
Philippines
Poland
Portugal
Qatar
Romania
Russia
Rwanda
Saudi Arabia
Senegal
Serbia
Sierra Leone
Singapore
Slovakia
Slovenia
Somaliland region
South Africa
South Korea
Spain
Sri Lanka
Sudan
Suriname
Swaziland
Sweden
Switzerland
Syria
Taiwan
Tajikistan
Tanzania
Thailand
Togo
Trinidad and Tobago
Tunisia
Turkey
Turkmenistan
Uganda
Ukraine
United Arab Emirates
United Kingdom
United States
Uruguay
Uzbekistan

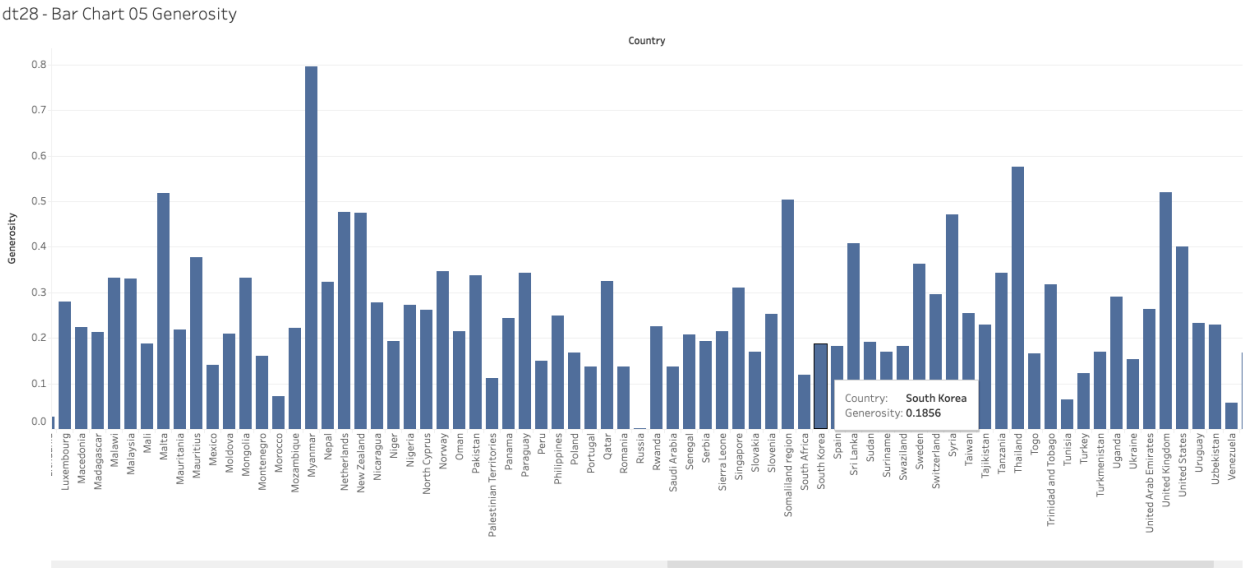
03.80


dt28 - Bar Chart 03 Happiness
Country
Happiness Score
7
6
5
4
3
2
1
0
Country: South Korea
Happiness Score: 5.984
Libya
Lithuania
Luxembourg
Macedonia
Madagascar
Malawi
Malaysia
Mali
Malta
Mauritania
Mauritius
Mexico
Moldova
Mongolia
Montenegro
Morocco
Mozambique
Myanmar
Nepal
Netherlands
New Zealand
Nicaragua
Niger
Nigeria
North Cyprus
Norway
Oman
Pakistan
Palestinian Territories
Panama
Paraguay
Peru
Philippines
Poland
Portugal
Qatar
Romania
Russia
Rwanda
Saudi Arabia
Senegal
Serbia
Sierra Leone
Singapore
Slovakia
Slovenia
Somaliland region
South Africa
South Korea
Spain
Sri Lanka
Sudan
Suriname
Swaziland
Sweden
Switzerland
Syria
Taiwan
Tajikistan
Tanzania
Thailand
Togo
Trinidad and Tobago
Tunisia
Turkey
Turkmenistan
Uganda
Ukraine
United Arab Emirates
United Kingdom
United States
Uruguay

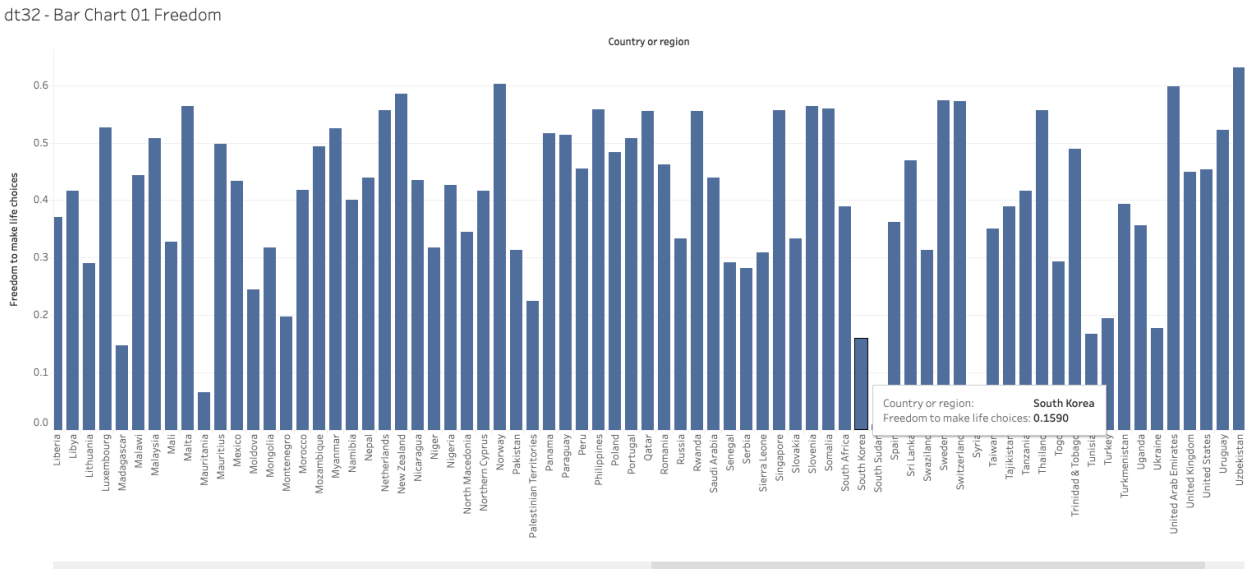
03.81


dt28 - Bar Chart 04 Gov't Trust
Country
Trust (Government Corruption)
0.0
0.1
0.2
0.3
0.4
0.5
Country: South Korea
Trust (Government Corruption): 0.0786

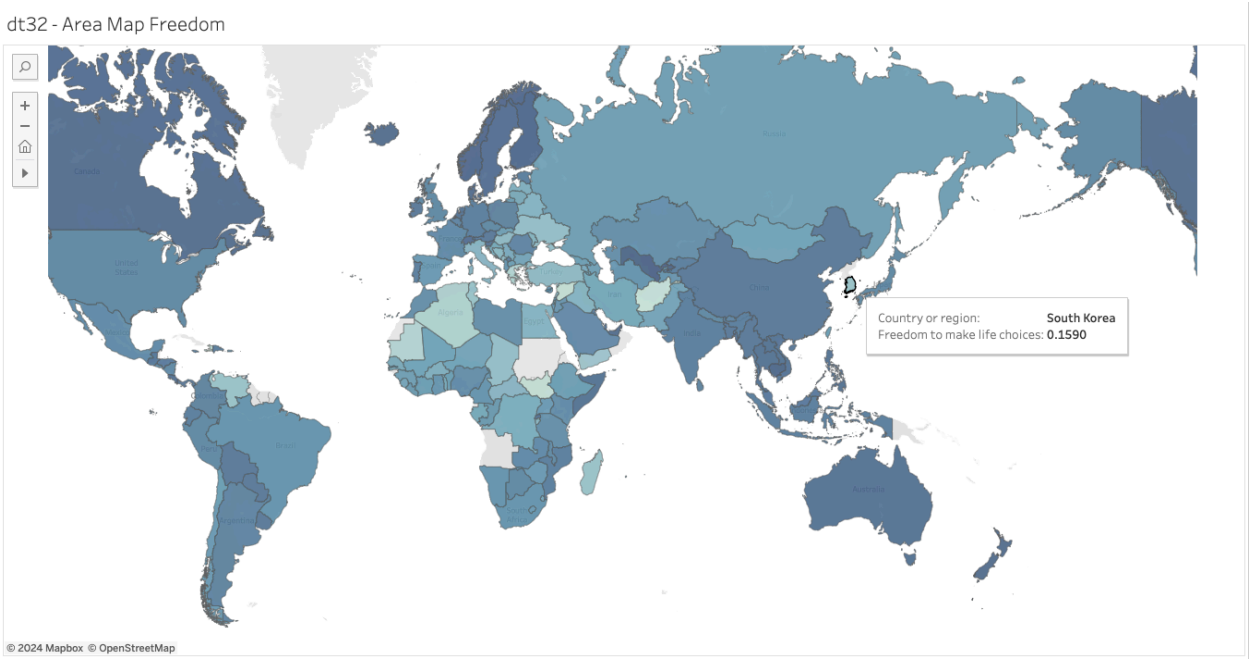
03.82


dt28 - Bar Chart 05 Generosity
Country
Generosity
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Country: South Korea
Generosity: 0.1856

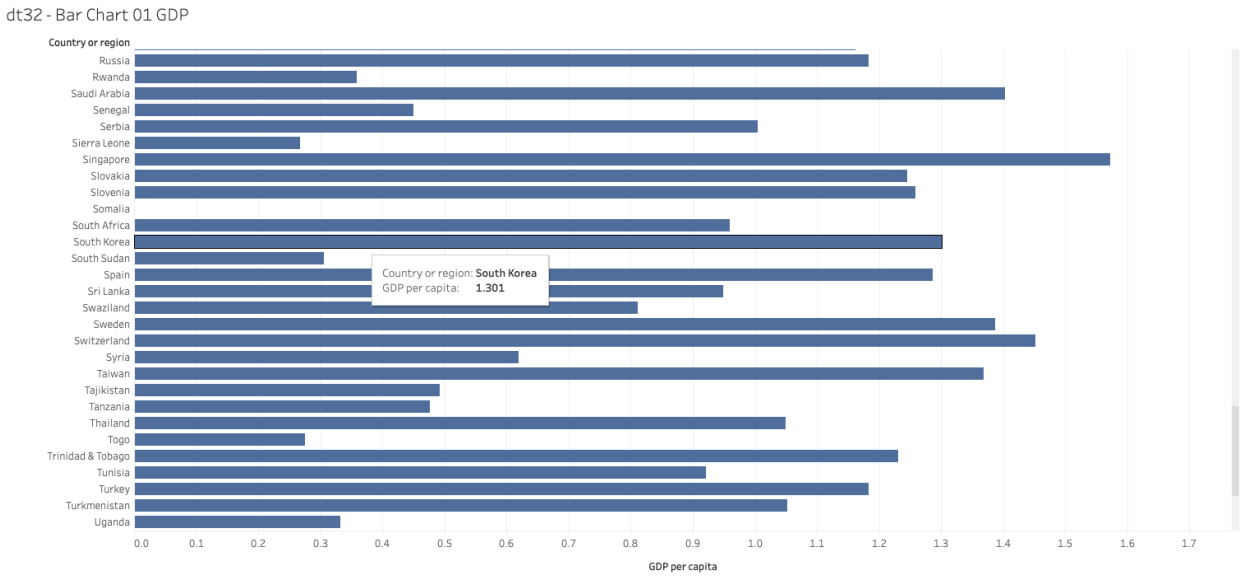
03.83


dt32 - Bar Chart 01 Freedom
Country or region
Freedom to make life choices
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Country or region: South Korea
Freedom to make life choices: 0.1590

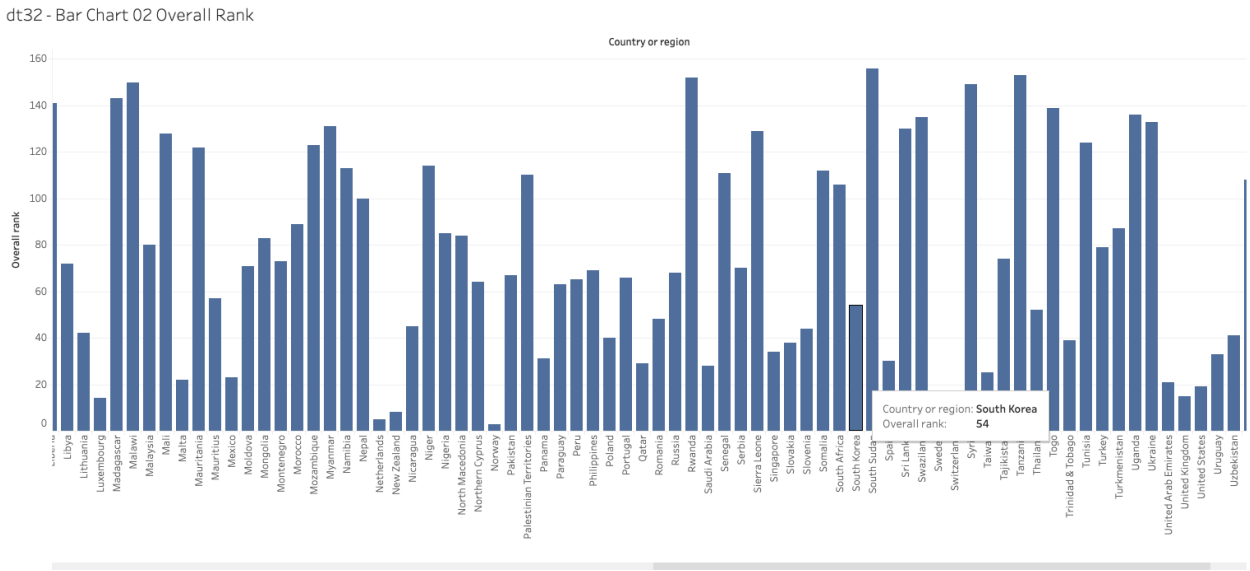
03.84


dt32 - Area Map Freedom
Country or region: South Korea
Freedom to make life choices: 0.1590
© 2024 Mapbox © OpenStreetMap

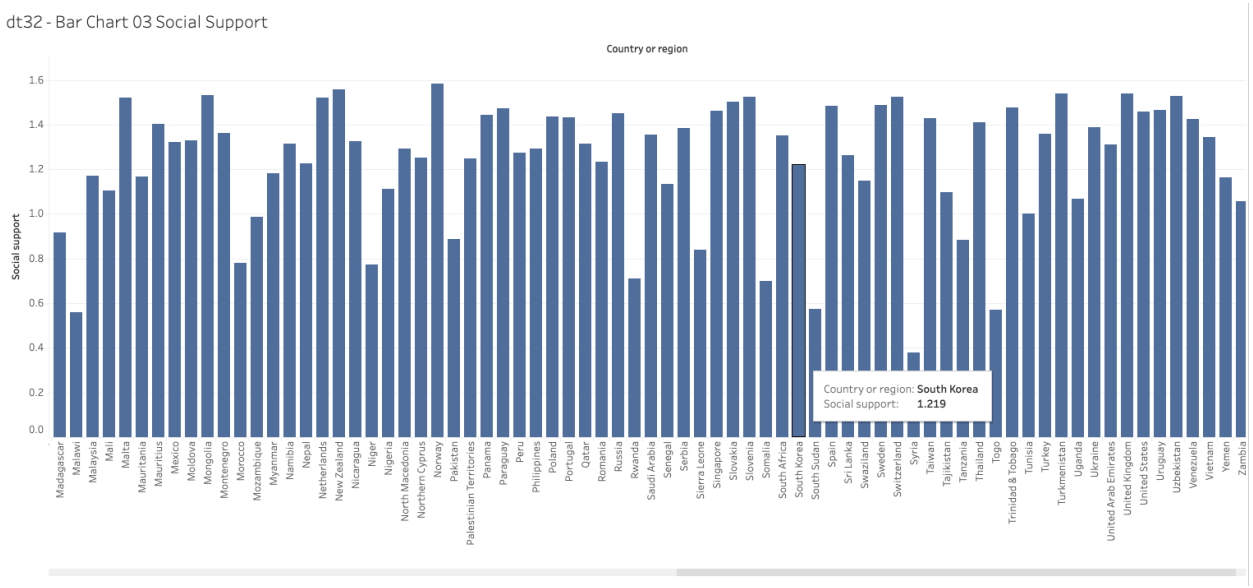
03.85


dt32 - Bar Chart 01 GDP
Country or region
Russia
Rwanda
Saudi Arabia
Senegal
Serbia
Sierra Leone
Singapore
Slovakia
Slovenia
Somalia
South Africa
South Korea
South Sudan
Spain
Sri Lanka
Swaziland
Sweden
Switzerland
Syria
Taiwan
Tajikistan
Tanzania
Thailand
Togo
Trinidad & Tobago
Tunisia
Turkey
Turkmenistan
Uganda
Country or region: South Korea
GDP per capita: 1.301
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7
GDP per capita

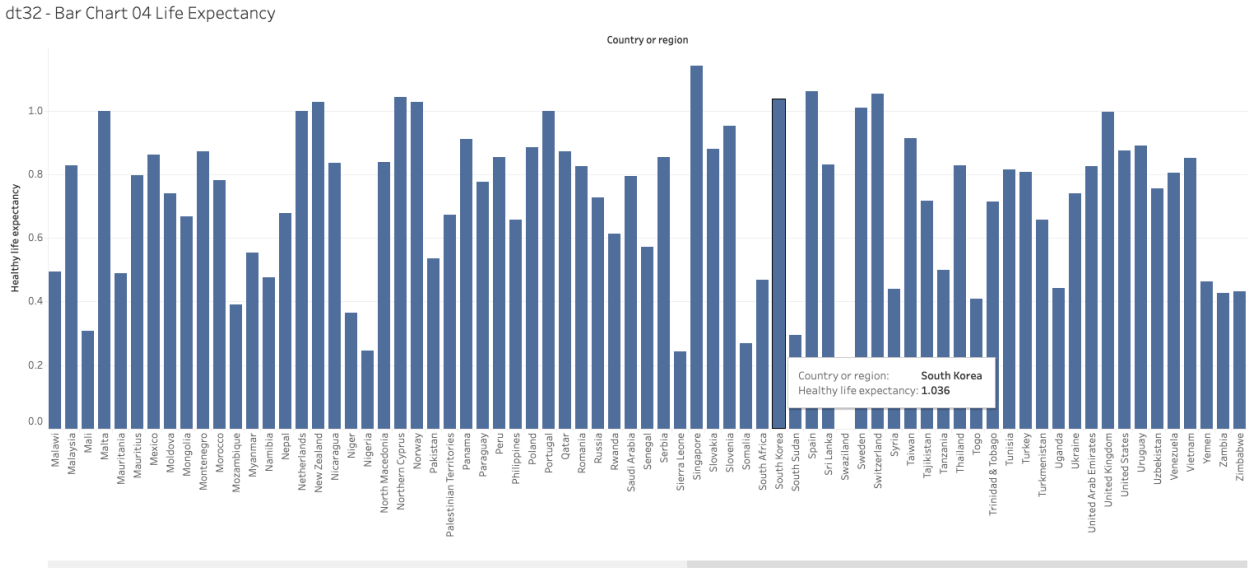
03.86


dt32 - Bar Chart 02 Overall Rank
Country or region
Overall rank
160 140 120 100 80 60 40 20 0
Country or region: South Korea
Overall rank: 54

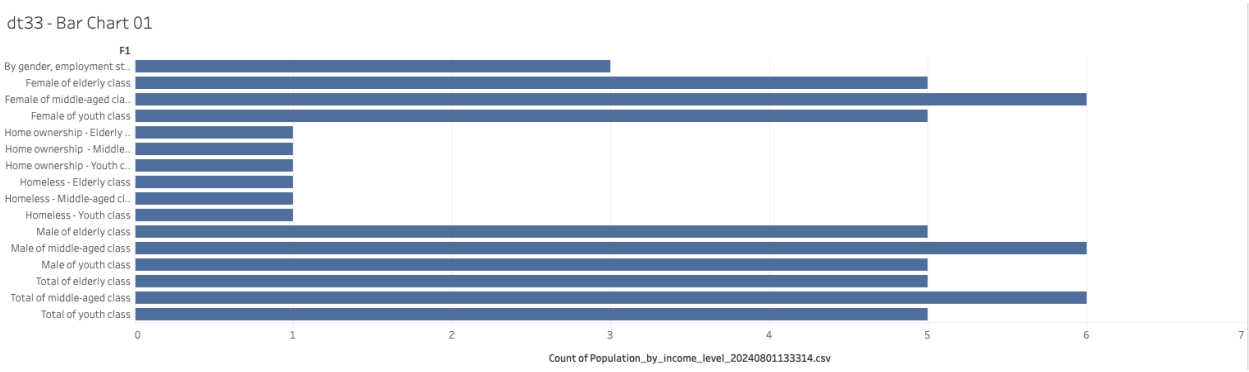
03.87


dt32 - Bar Chart 03 Social Support
Country or region
Social support
Country or region: South Korea
Social support: 1.219

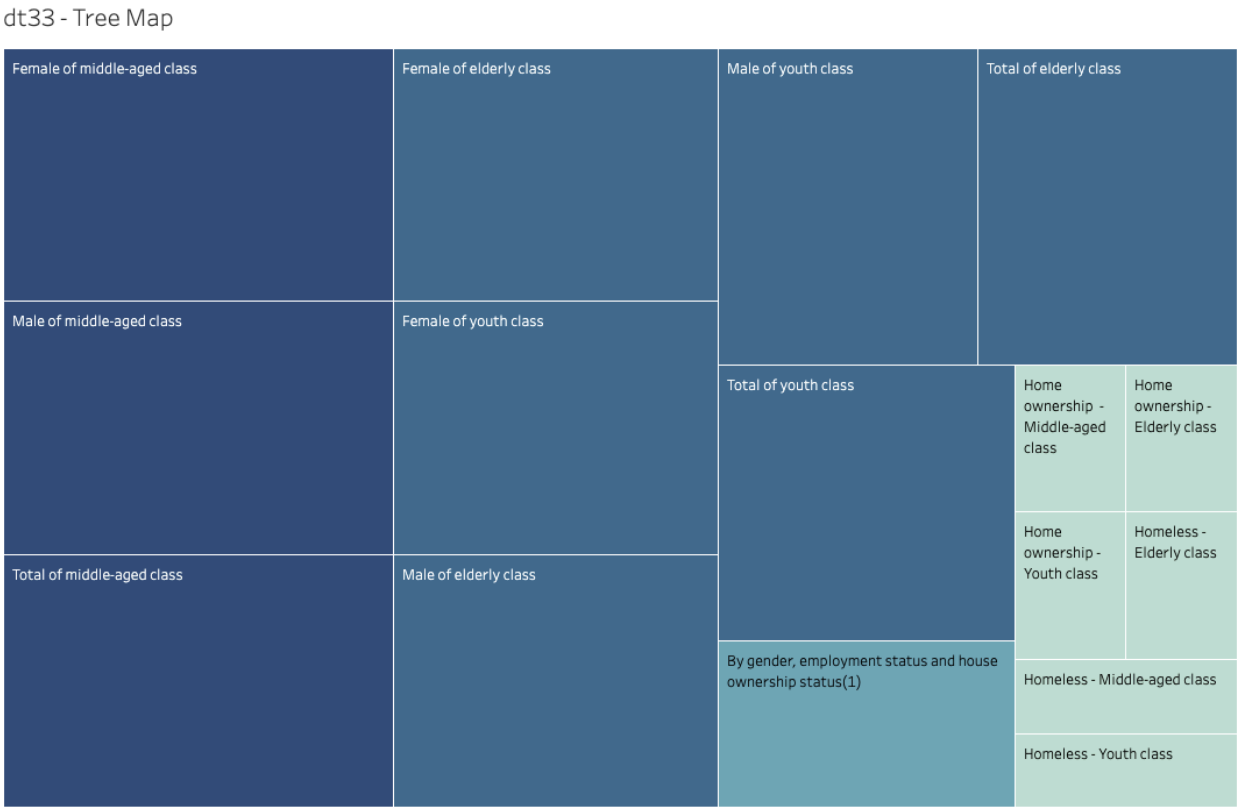
03.88


dt32 - Bar Chart 04 Life Expectancy
Country or region
Healthy life expectancy
Country or region: South Korea
Healthy life expectancy: 1.036

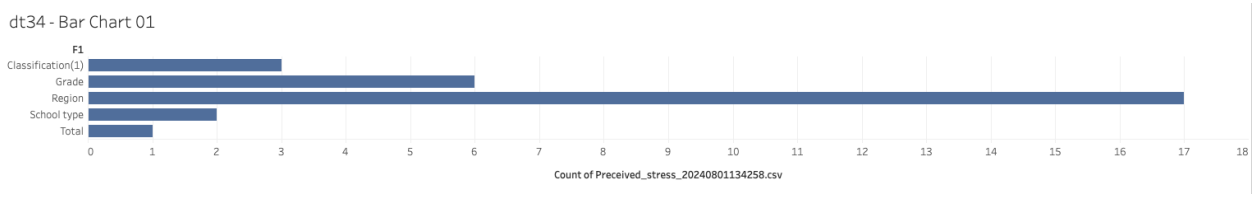
03.89


dt33 - Bar Chart 01
F1
By gender, employment st..
Female of elderly class
Female of middle-aged cla..
Female of youth class
Home ownership - Elderly ..
Home ownership - Middle..
Home ownership - Youth c..
Homeless - Elderly class
Homeless - Middle-aged cl..
Homeless - Youth class
Male of elderly class
Male of middle-aged class
Male of youth class
Total of elderly class
Total of middle-aged class
Total of youth class
0
1
2
3
4
5
6
7
Count of Population_by_income_level_20240801133314.csv

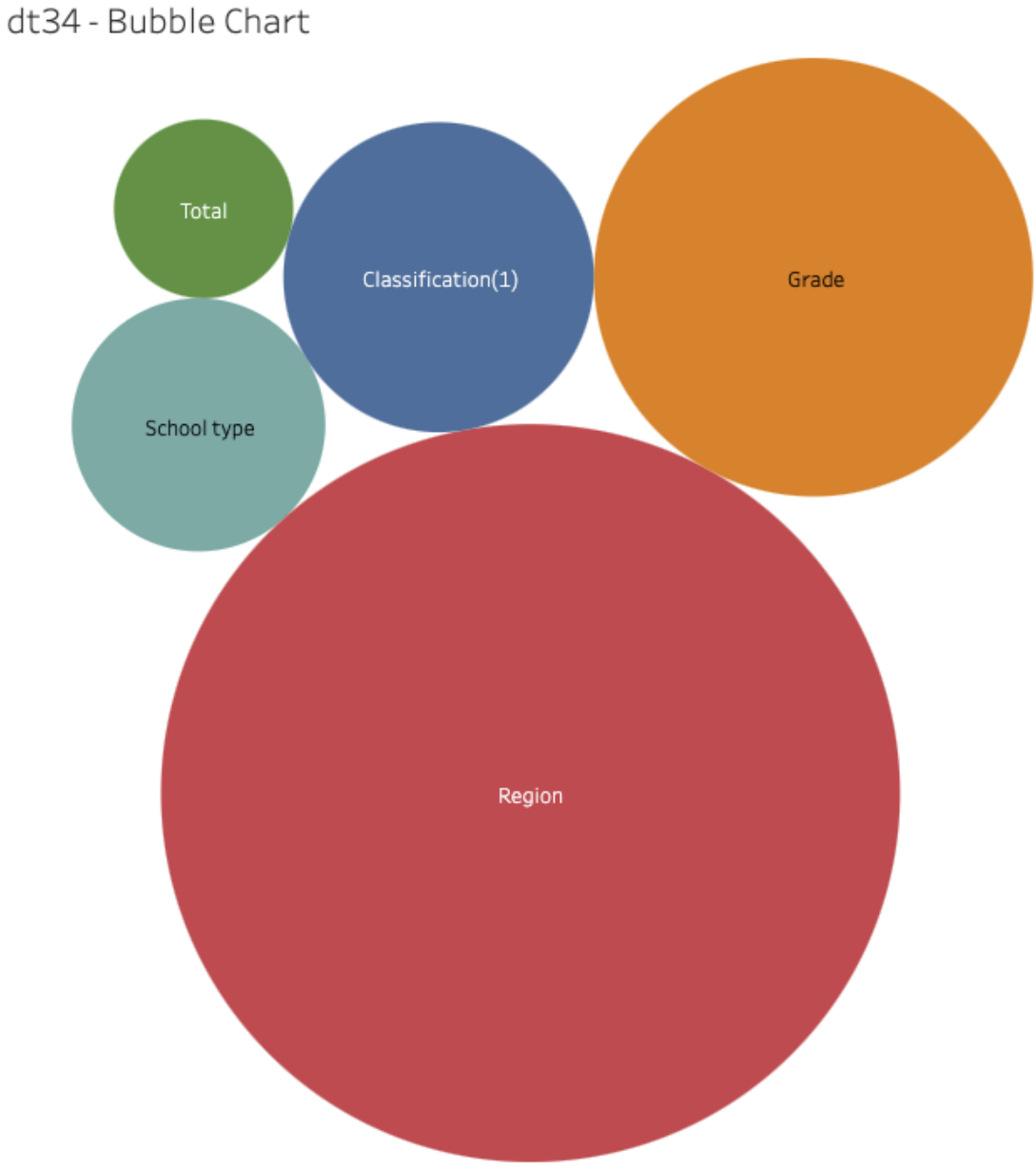
03.90


dt33 - Tree Map
Female of middle-aged class
Female of elderly class
Male of youth class
Total of elderly class
Male of middle-aged class
Female of youth class
Total of youth class
Home ownership - Middle-aged class
Home ownership - Elderly class
Home ownership - Youth class
Homeless - Elderly class
Total of middle-aged class
Male of elderly class
By gender, employment status and house ownership status(1)
Homeless - Middle-aged class
Homeless - Youth class

03.91

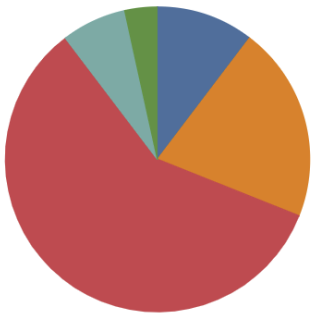

dt34 - Bar Chart 01
F1
Classification(1)
Grade
Region
School type
Total
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Count of Preceived_stress_20240801134258.csv

03.92


dt34 - Bubble Chart
Total
Classification(1)
Grade
School type
Region

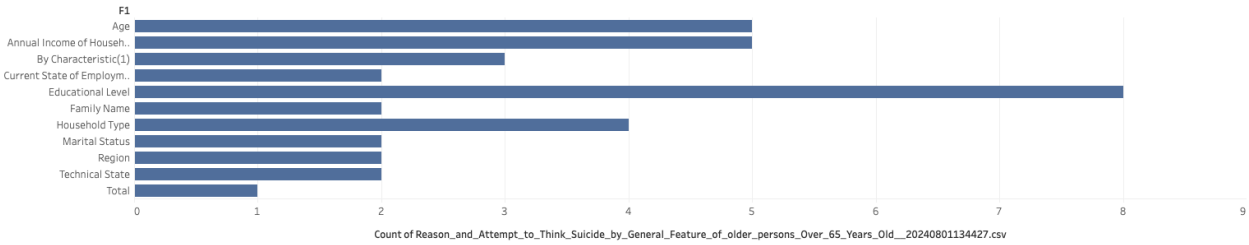
03.93

dt34 - Pie Chart



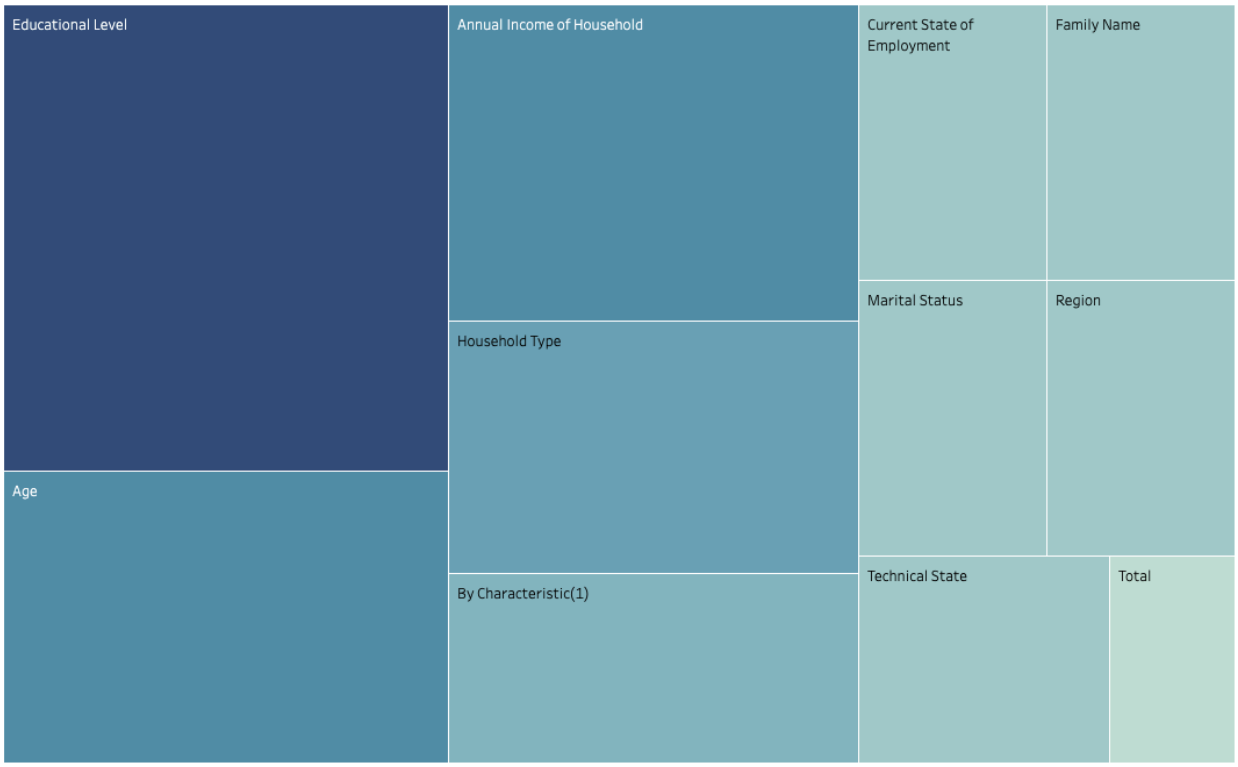
03.94

dt35 - Bar Chart



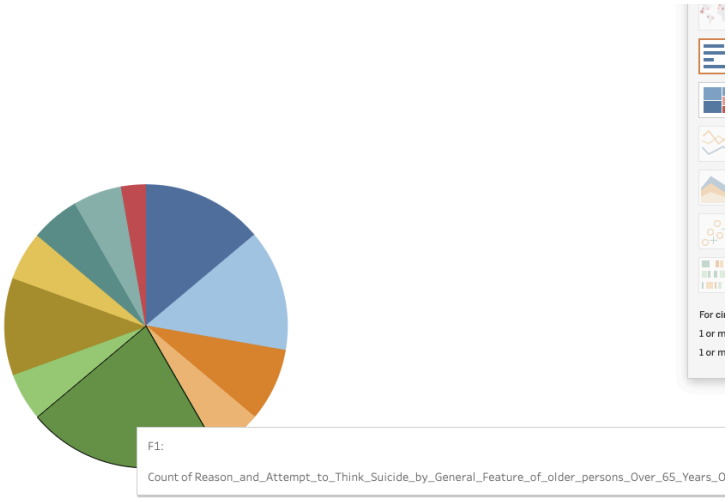
03.95

dt35 - Tree Map

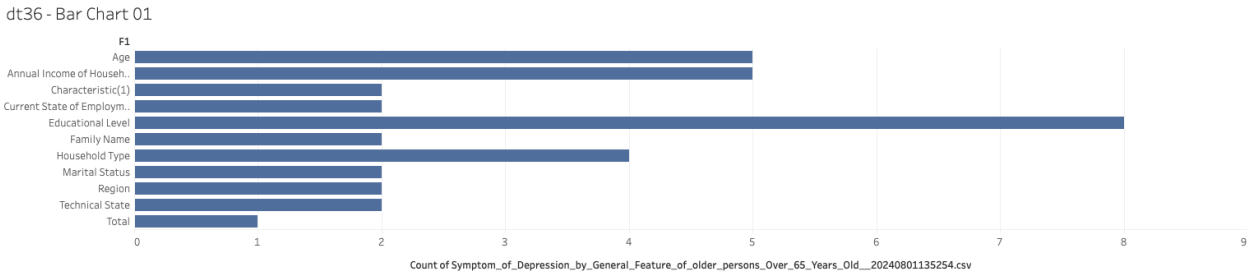


03.96

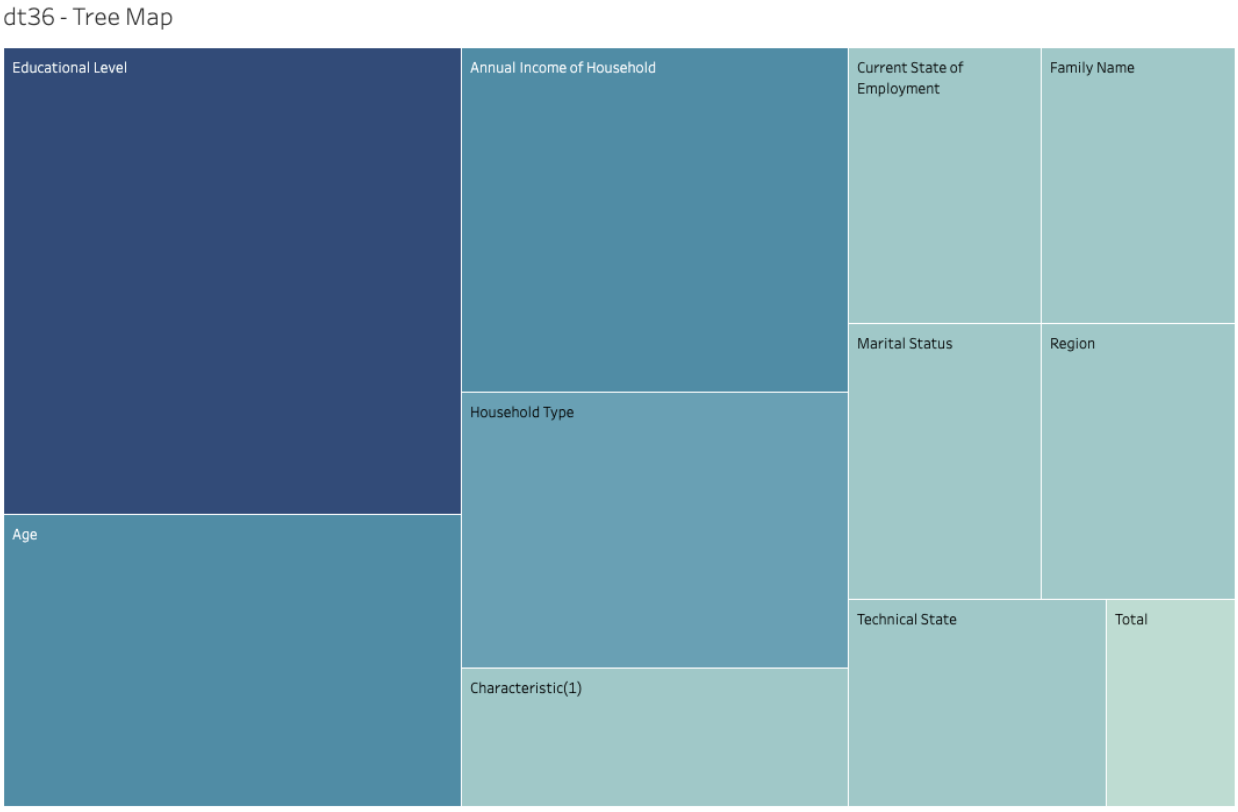
dt35 - Pie Chart



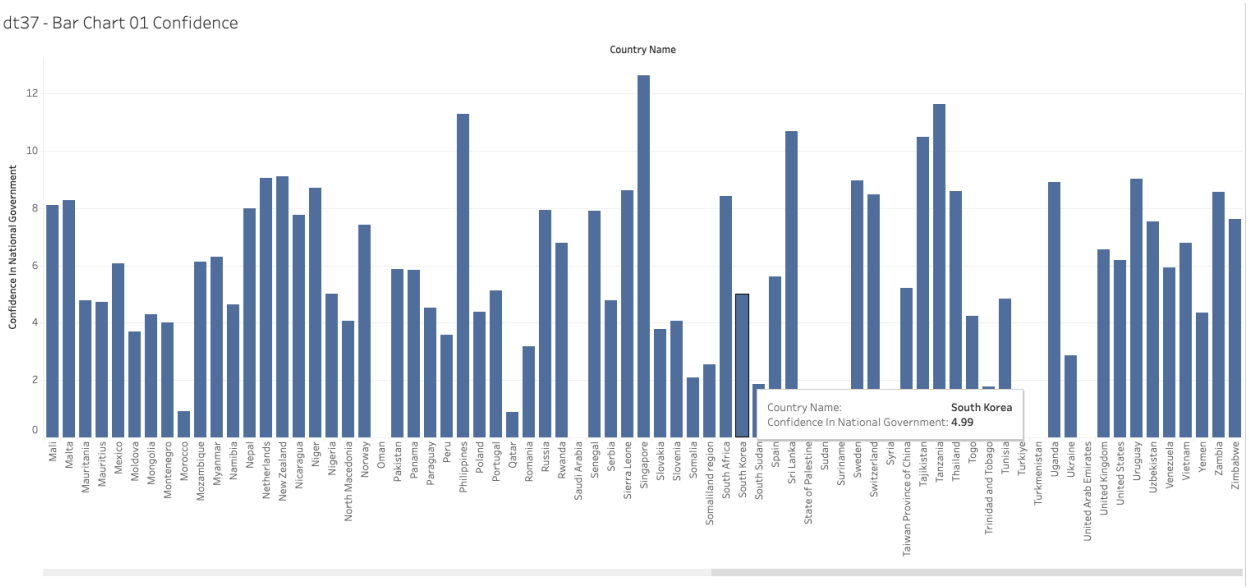
03.97


dt36 - Bar Chart 01
F1
Age
Annual Income of Househ..
Characteristic(1)
Current State of Employm..
Educational Level
Family Name
Household Type
Marital Status
Region
Technical State
Total
0
1
2
3
4
5
6
7
8
9
Count of Symptom_of_Depression_by_General_Feature_of_older_persons_Over_65_Years_Old__20240801135254.csv

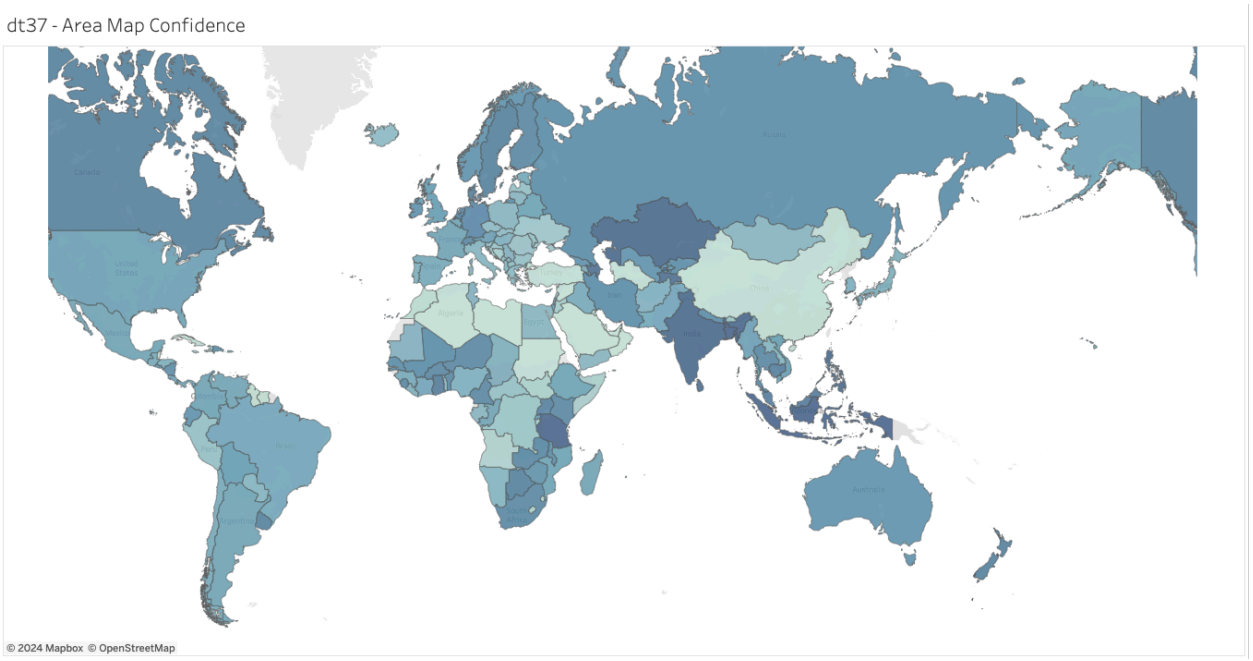
03.98


dt36 - Tree Map
Educational Level
Annual Income of Household
Current State of Employment
Family Name
Marital Status
Region
Household Type
Age
Technical State
Total
Characteristic(1)

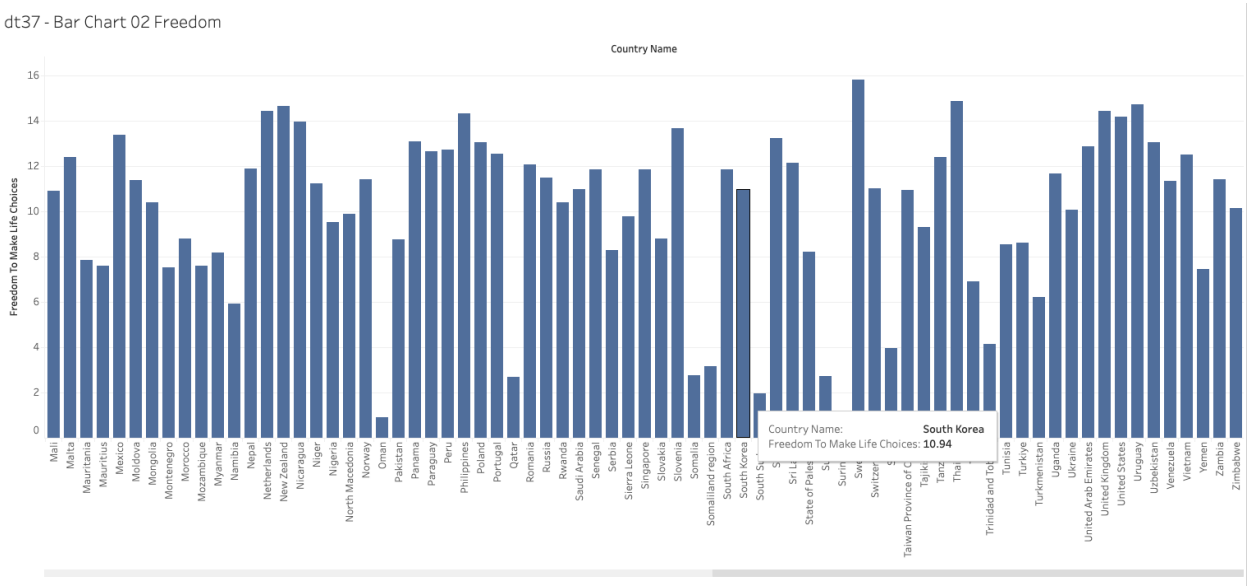
03.99


dt37 - Bar Chart 01 Confidence
Country Name
Confidence In National Government
0
2
4
6
8
10
12
Country Name: South Korea
Confidence In National Government: 4.99
Mali
Malta
Mauritania
Mauritius
Mexico
Moldova
Mongolia
Montenegro
Morocco
Mozambique
Myanmar
Namibia
Nepal
Netherlands
New Zealand
Nicaragua
Niger
Nigeria
North Macedonia
Norway
Oman
Pakistan
Panama
Paraguay
Peru
Philippines
Poland
Portugal
Qatar
Romania
Russia
Rwanda
Saudi Arabia
Senegal
Serbia
Sierra Leone
Singapore
Slovakia
Slovenia
Somalia
Somaliland region
South Africa
South Korea
South Sudan
Spain
Sri Lanka
State of Palestine
Sudan
Suriname
Sweden
Switzerland
Syria
Taiwan Province of China
Tajikistan
Tanzania
Thailand
Togo
Trinidad and Tobago
Tunisia
Türkiye
Turkmenistan
Uganda
Ukraine
United Arab Emirates
United Kingdom
United States
Uruguay
Uzbekistan
Venezuela
Vietnam
Yemen
Zambia
Zimbabwe

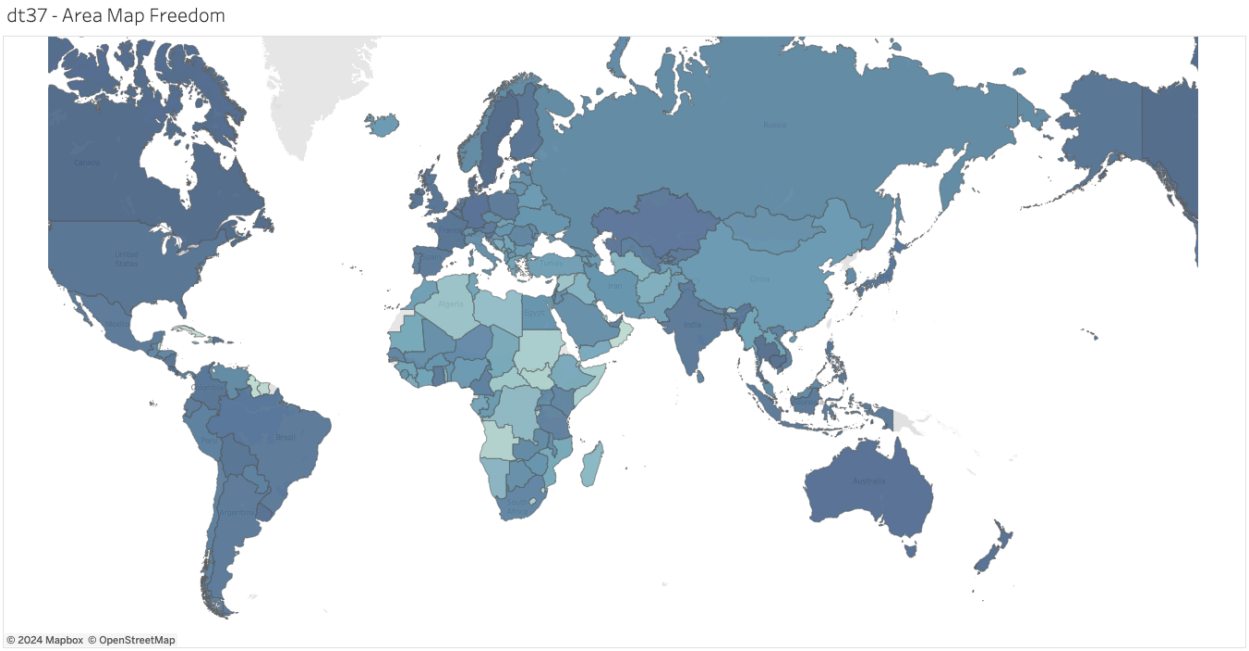
03.100


dt37 - Area Map Confidence
© 2024 Mapbox © OpenStreetMap

03.101

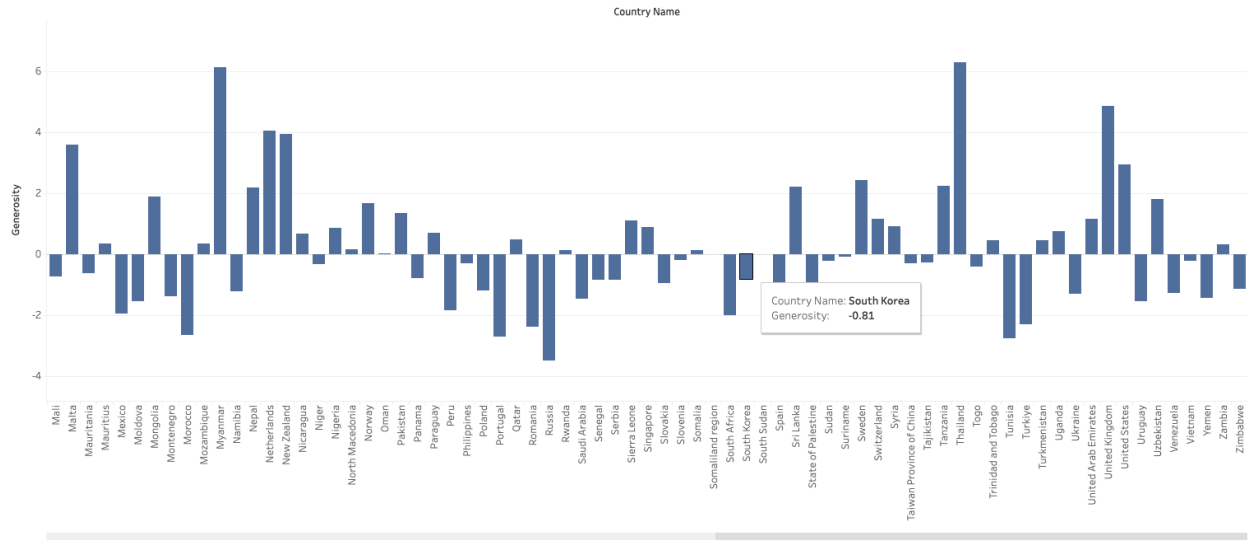

dt37 - Bar Chart 02 Freedom
Country Name
Freedom To Make Life Choices
Country Name: South Korea
Freedom To Make Life Choices: 10.94

03.102


dt37 - Area Map Freedom
© 2024 Mapbox © OpenStreetMap

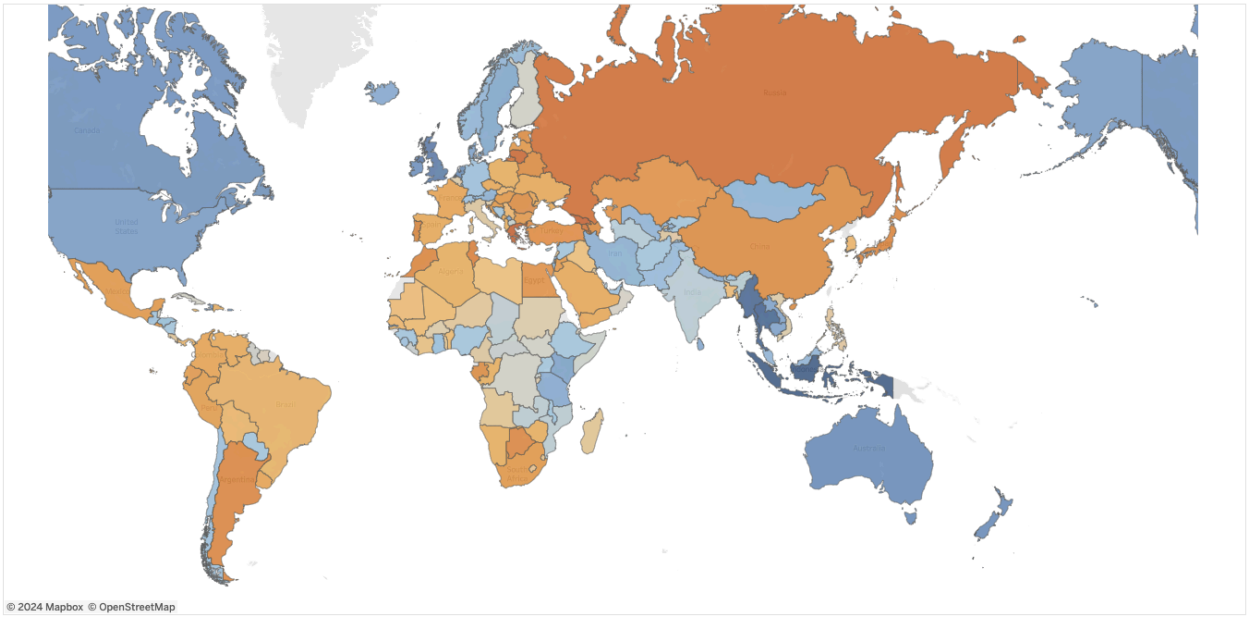
03.103

dt37 - Bar Chart 03 Generosity



03.104

dt37 - Area Map Generosity



Dataset 01:

‘101 _DT _1B34E09 _20240718123555.csv’ as dt01 _death _cause _gend _____00

Dataset 02:

‘101 _DT _1B34E12 _20240718123805.csv’ as ‘dt02 _death _cause _geo _____00’

Dataset 03:

‘who_suicide_statistics.csv’ as ‘dt03_who_suicide_____00’

Dataset 04:

‘Combined_processed_data.csv’ as ‘dt04_combined_____00’

Dataset 05:

'Degree_of_Stress__General_Life__13_years_old_and_over__20240719092712.csv' as ‘dt05_stress_general_____00’

Dataset 06:

'Degree_of_Stress__Home_Life__13_years_old_and_over__2024071909291 4.csv' as ‘dt06_stress_home_____00’

Dataset 07:

'Degree_of_Stress__School_Life__13_years_old_and_over__20240719092757.csv' as ‘dt07_stress_school_____00’

Dataset 08:

'Degree_of_Stress__Work_Life__13_years_old_and_over__20240719092838.csv' as ‘dt08_stress_work_____00’

Dataset 09:

'Drinking__19_years_old_and_over__20240719093320.csv' as ‘dt09_drinking_19_____00’

Dataset 10:

'Drinking__20_years_old_and_over__20240719093241.csv' as ‘dt10_drinking_20_____00’

Dataset 11:

'Drinking_and_Health_Management__19_years_old_and_over__20240719093528.csv' as ‘dt11_drinking_manage_19_____00’

Dataset 12:

'Drinking_and_Health_Management__20_years_old_and_over__20240719093453.csv' as ‘dt12_drinking_manage_20_____00’

Dataset 13:

'Impulse_to_Commit_Suicide_and_Reasons__13_years_old_and_over__20240719092337.csv' as ‘dt13_suicide_impulse_____00’

Dataset 14:

'Impulse_to_Commit_Suicide_and_Reasons__13_years_old_and_over__20240719092337.csv' as ‘dt14_suicide_impulse_____00’

Dataset 15:

'Reason_and_Attempt_to_Think_Suicide_by_General_Feature_of_older_persons_Over_65_Years_Old__20240719092517.csv' as ‘dt15_suicide_reason_____00’

Dataset 16:

'Reason_and_Attempt_to_Think_Suicide_by_General_Feature_of_older_persons_Over_65_Years_Old__20240719092517.csv' as ‘dt16_suicide_reason_____00’

Dataset 17:

'Smoking_and_Drinking__19_years_old_and_over__20240719093138.csv' as ‘dt17_smoke_drink_19_____00’

Dataset 18:

'Smoking_and_Drinking__20_years_old_and_over__20240719093056.csv' as ‘dt18_smoke_drink_20_____00’

Dataset 19:

'Ph_categories_index.csv' as ‘dt19_ph_categories_____00’

Dataset 20:

'ph_Pornhub Analysis year by year.xlsx' as ‘dt20_ph_analysis_____00’

Dataset 21:

'Ph_videos.csv' as ‘dt21_ph_videos_____00’

Dataset 22:

‘pd.concat(chunk)’ as ‘dt22_ph_tot_____00’

Dataset 23:

'408_DT_40803_N0003_20240801134720.csv' as

‘dt23_408_03_____00’

Dataset 24:

'408_DT_40803_N0004_20240801134840.csv' as

‘dt24_408_04_____00’

Dataset 25:

'Economic_Sentiment_Index_20240801135039.csv' as

‘dt25_index_eco_sent_____00’

Dataset 26:

'Feeling_sad_or_hopeless_20240801134129.csv' as

‘dt26_sadness_____00’

Dataset 27:

'Happiness.csv' as ‘dt27_happiness_01_____00’

Dataset 28:

'index_happiness_2015.csv' as ‘dt28_happiness_2015_____00’

Dataset 29:

'index_happiness_2016.csv' as ‘dt29_happiness_2016_____00’

Dataset 30:

'index_happiness_2017.csv' as ‘dt30_happiness_2017_____00’

Dataset 31:

'Index_happiness_2018.csv' as ‘dt31_happiness_2018_____00’

Dataset 32:

'index_happiness_2019.csv' as ‘dt32_happiness_2019_____00’

Dataset 33:

'Population_by_income_level_20240801133314.csv' as

‘dt33_pop_income_____00’

Dataset 34:

'Preceived_stress_20240801134258.csv' as

‘dt34_stress_perc_____00’

Dataset 35:

'Reason_and_Attempt_to_Think_Suicide_by_General_Feature_of_older_persons_Over_65_Years_Old__20240801134427.csv' as ‘dt35_suic_reason_____00’

Dataset 36:

'Symptom_of_Depression_by_General_Feature_of_older_persons_Over_65_Years_Old__20240801135254.csv' as ‘dt36_depr_symptom_____00’

Dataset 37:

'World Happiness Report.csv' as ‘dt37_happiness_world_____00’

Ten questions an audience would ask you:

1. What are some political factors that had significance in your research?
    a. Unfortunately, I was unable to delve into the political factors that could affect my results. I do, however, highly doubt that political factors have a significant effect on the suicide rate.
2. What demographic factors had significance in your research?
    a. There are significant demographic factors that affect the suicide rate. The main culprit is the generational factor. My research returned indisputable evidence that the age group of above 64 years of age had a disproportionately higher rate of suicide.
3. What are some social aspects that had significance in your research?
    a. Unfortunately, I was unable to delve into the social factors that could affect my results. I do, however, think that LGBT intolerance in the ROK has an effect on the suicide rate.
4. What economic factors had significance in your research?
    a. There is a correlation with the state of the economy and the suicide rate. This is based on historical data which showed a correlation between the economic situation in the early 1980s and fluctuation in suicides at the time.

5. Are there any environmental considerations that should be included in your research?
    a. I don’t believe that environmental factors should be taken into consideration regarding this topic. The environment of the ROK does not reach extremes that would affect the general population.
6. Are there any genetic considerations that should be included in your research?
    a. Unfortunately, I was unable to conduct research regarding this issue. This type of research would require vast amounts of data across all demographics and including scientific, particularly biological data.
7. How much does substance abuse affect the results of your models?
    a. I analyzed several datasets containing statistics on substance abuse in the ROK. I was unable to find any significant correlations with substance abuse and the suicide rate.
8. Does happiness index correlate with your results?
    a. When making comparisons to other countries, there is a slight correlation with the happiness index and the suicide rate. Since the metrics used are undetermined in the data, I would need to delve further into what metrics are used.

9. What metrics are used to measure happiness index?
    a. The datasets that I obtained did not elaborate on the metrics used to calculate the happiness index. I would need to conduct further research into this.
10. Based on your results, how can you explain the gender gap regarding the issue?
    a. It’s obvious, based on my results, that males have a much higher rate of suicide than females. I wasn’t able to uncover the reasons for this due to the lack of relevant data. Despite males more than doubling the number of suicides of females, the number of attempted suicides are comparable between the two.