Milestone Three

Ross L. Kim-Schreck

DSC680 Applied Data Science

Professor Iranitalab

2024.06.30

Exploring Policies to Curb the Rate of Stomach Cancer in the ROK

Business Problem

In this case, it is more of a social problem. The fatality rate of men in the Republic of Korea, which I will refer to as 'the ROK', is far higher than their female counterparts. The main culprits are diseases such as stomach cancer, lung cancer, and liver cancer. There are a number of factors affecting this. The gender gap is still an issue in Korea. The number of working men far outnumbers that of women. Korean society is still conservative, so women are typically homemakers while men have careers. Due to Korea's corporate culture, drinking alcohol is very common among office workers. In many companies drinking is compulsory where office dinners are held every week or so. Because of this, many men also smoke tobacco or vape. The significantly lower number of women who drink and smoke is indicative of the large gender gap.

Background / History

Korean corporate culture has been one based on strict hierarchy since the end of the Korean war. This is in a major part due to the influence of confucianism which originated in neighboring China. Every facet of life in Korea is affected by this, from personal interactions to entire industries. The corporate culture of Korea is no exception. Despite being an impoverished war-torn nation post Korean war era, South Korea is a flourishing democracy with the 13th largest economy in the world and third largest in Asia. In the

past, Seoul's infrastructure was not developed enough to accommodate a high population, so the South remained mostly agrarian for a considerable span of time. The government, which was a military dictatorship, decided to rebuild and revive itself after the Korean war by taking an 'economy first' initiative. While technically a democracy, the quasi dictatorship hand picked certain families to found conglomerates (Chaebols) which were then assigned specific industries in which they could maneuver, simply speaking, every South Korean conglomerate, to the present day, is constructed and functions like a monarchy inside this pseudo democracy – the ROK – each 'monarchy' has the CEO as king; this is nepotism and monarchism existing in a modern-day liberal democracy. The South Korean conglomerate is not so different from the government of North Korea. The South Korean economy thrives on one of its only resources: human resources. These resources are expected life-long commitments to their employers. They are also expected to work 10-12 hour days and often weekends when work loads are larger. South Korea ranked among the lowest in overall happiness among OECD nations; Korea's corporate culture is the main culprit for this. This is a major source of stress on the individual which, in turn, leads to binge drinking. Thus, at an individual level, it is common for laborers to strive to get into the largest conglomerates which is key, in their eyes, to life-long success and stability. This has spawned a strict drinking culture in which it is mandatory and considered of utmost importance in corporate Korea.

Data Explanation

From milestone one, I initially used two datasets from KOSIS (Korean Statistical Information Service), which is the Korean government's official website: 'Statistics on Stomach Cancer Screening Diagnosis by Region and Gender' and 'Cancer Incident Cases and Incidence Rates by Site and Sex'. The first dataset contains statistics on stomach cancer in particular, the dataset only contains data for new cases in 2021, and the second dataset contains statistics on common cancer diagnoses throughout the ROK, the dataset only contains data for new cases in 2021. These datasets are inadequate for the type of research I want to conduct. *(Appendix 01.01) (Appendix 01.02)*

From milestone two, I retrieved three additional datasets: 'Worldwide Cancer Dataset (females)', 'Worldwide Cancer Dataset (males)', and 'Worldwide Cancer Dataset (total)'. Unfortunately, this dataset only contained data for new cases. *(Appendix 01.03)*

I then retrieved the 'Cancer Rates' dataset, but it only contained a handful of cancer types. *(Appendix 01.04) (Appendix 01.05)*

I then retrieved 'Cancer Deaths by Country and Type Dataset'. This dataset contains data on apparently every country and a multitude of cancer types. *(Appendix 01.06)*

For milestone three, I retrieved 'Gender Inequality Index.csv'. This dataset contains data on gender equality of all nations. I also retrieved

'Avg_hours_worked_(1950-2017).csv, which contains data of every country on average labor hours. I retrieved two additional datasets which are similar: 'Time_use_OECD.csv', which contains data on daily hours worked per nation and gender and 'Yearly_RGDPO_(1950-2017).csv', which contains data on the real GDP of all nations. *(Appendix 03.09)*

Methods

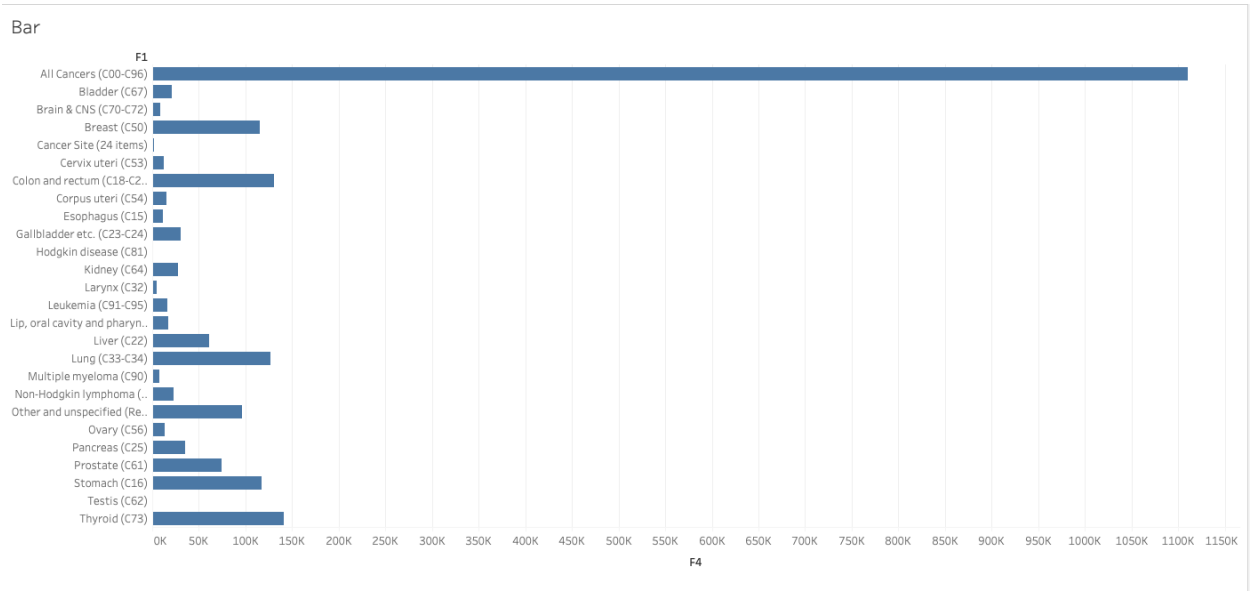
From milestone two, I tested the accuracy of my data using linear regression models. The data is a good fit for this model because it spans a time period of over 30 years. Using 'Stomach Cancer' as the target, I ran a regression model on all countries. It returned a very high RMSE of over 500 and R2 of 0.99. These figures are illogical. I then narrowed my data to South Korea and ran a model with the target being 'Stomach Cancer'. It returned an RMSE of 53.72 and R2 of 0.99, which is a better result but still way too high. I will transform the data accordingly by adjusting or removing outliers. I plan to run additional models on this dataset in the next milestone.

For milestone three, I tested the accuracy of 'Stomach Cancer' after making some transformations. It returned a lower RMSE and higher R2, but both are still too high. I think this is due to numerous outliers which would take an extended period of time to resolve. I also ran regression models on the additional datasets added for this milestone. Due to the high disparity

among nations and the nature of the data, accuracy was very hard to calculate. I would need a lot more time to return conclusive results.

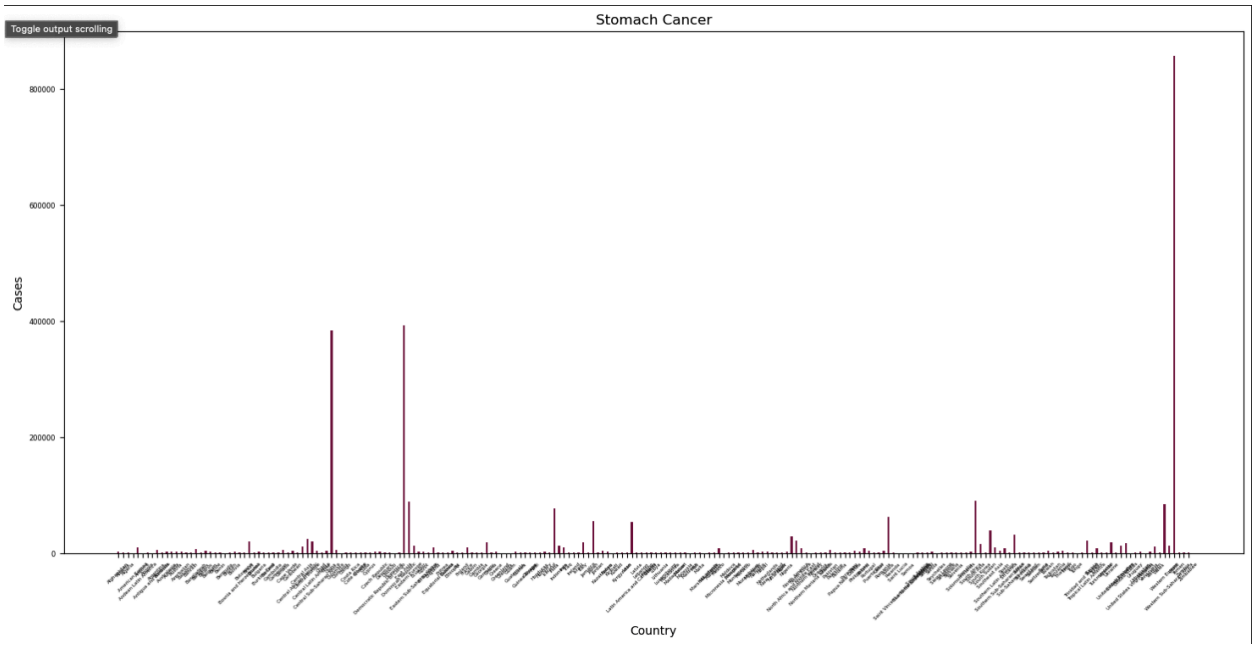
Analysis

Based on the last dataset, lung, breast, and thyroid cancers are among the most common in the world. *(Appendix 02.01)*

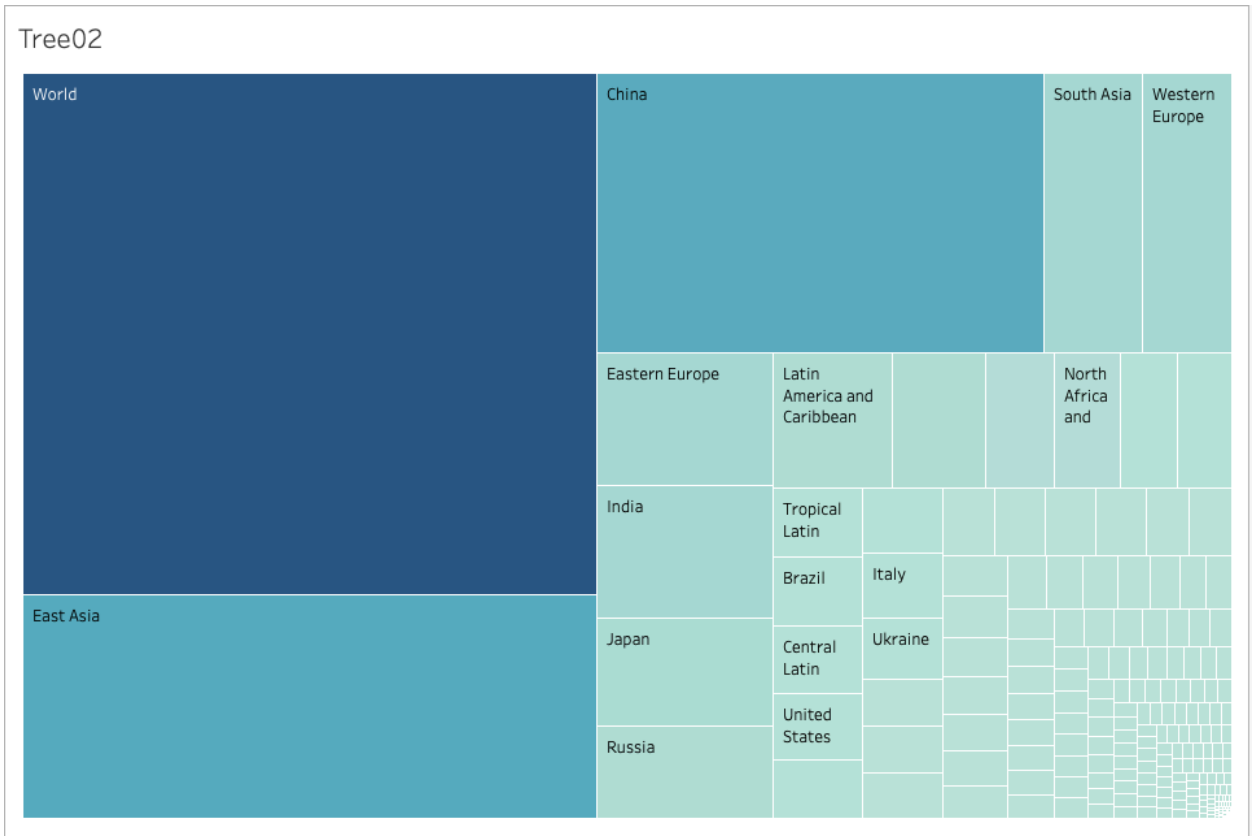


Based on this dataset, East Asian nations have the highest rate of death caused by stomach and liver cancer. There isn't enough room for all country names in the x-axis. I did confirm, however, that the East Asia region is the highest. *(Appendix 02.02) (Appendix 02.06) (Appendix 02.07)*

(Appendix 02.02)

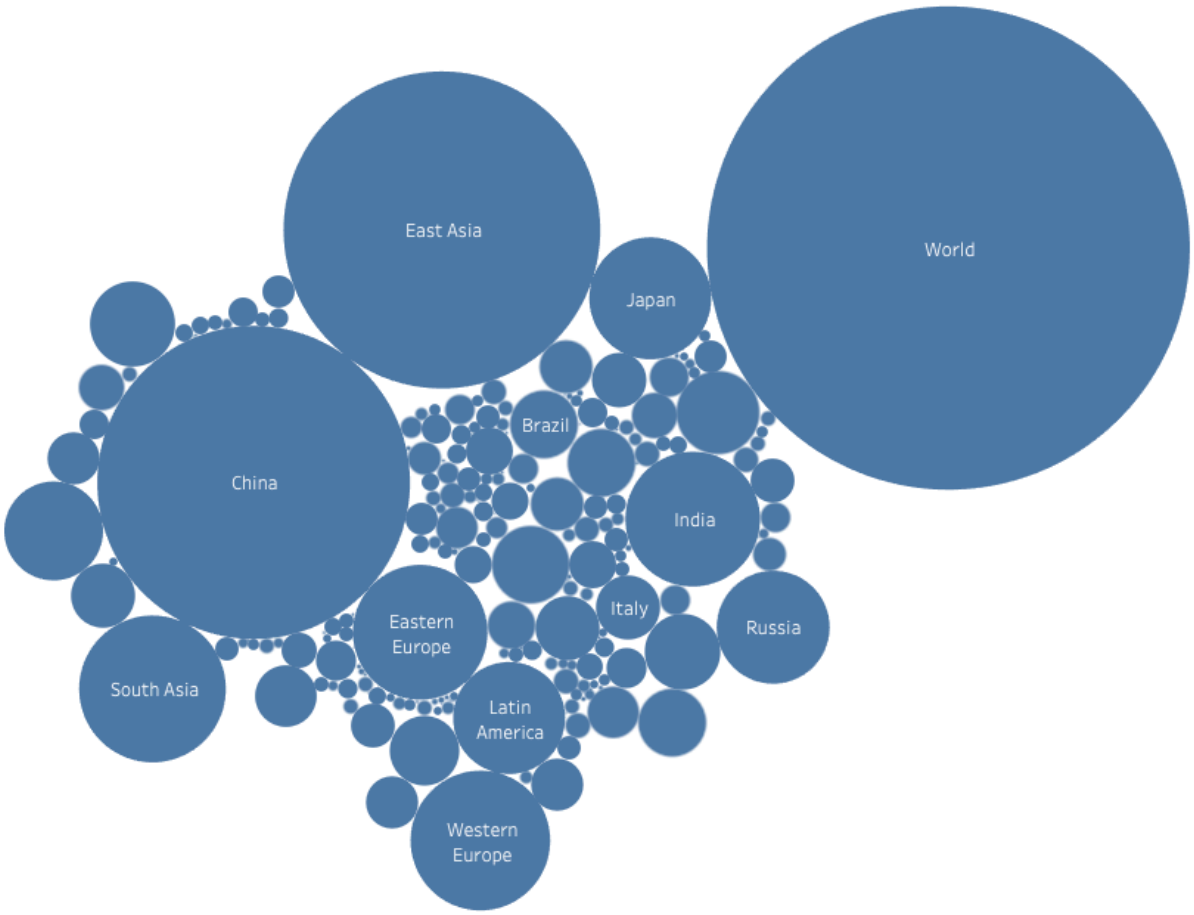


(Appendix 02.06)

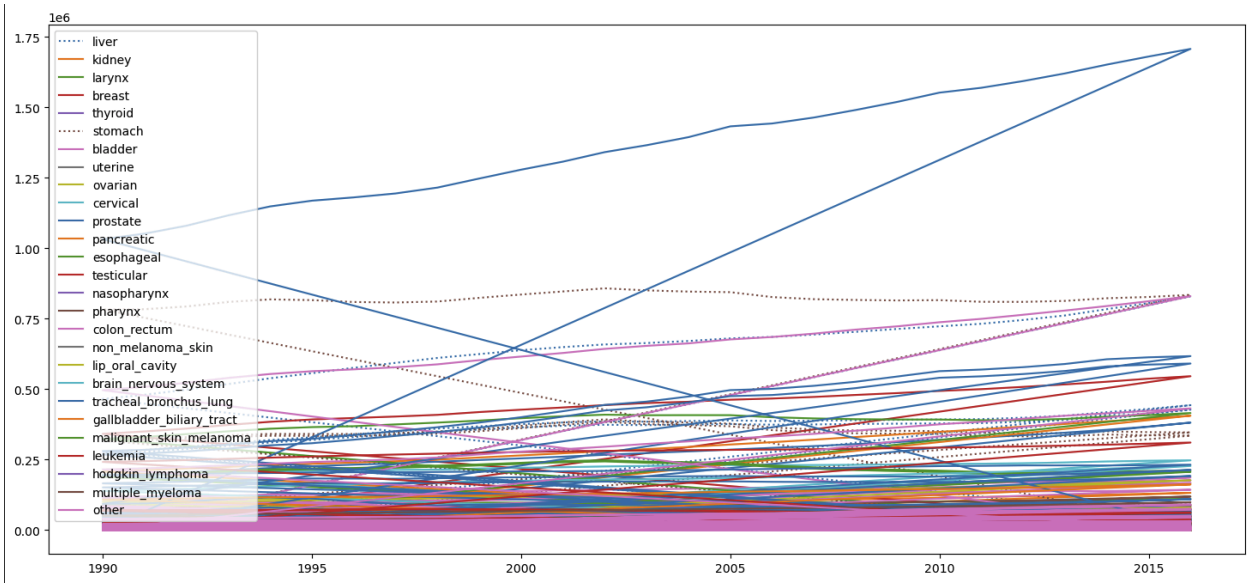


(Appendix 02.07)

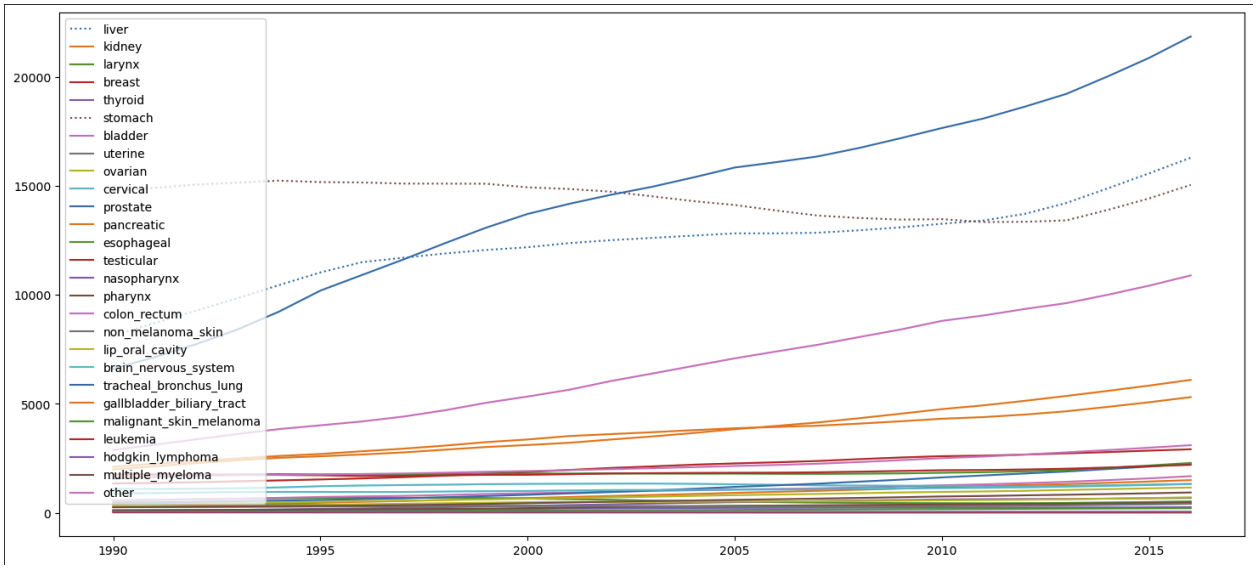
Tree02



Based on this dataset, liver and stomach cancers are not the highest among all nations. *(Appendix 02.03)*



Based on the same dataset, in the case of South Korea, stomach and liver cancers are among the highest. *(Appendix 02.04)*



Based on the same dataset, the most common cancers are brain, stomach, and lung. *(Appendix 03.01)*

Based on labor statistics dataset, Korea ranks the highest number of hours worked per day. *(Appendix 03.03) (Appendix 03.06)*

Based on the RGDPO (real gross domestic product) dataset, the US, China, and Japan rank the highest; the ROK is 11th. *(Appendix 03.07) (Appendix 03.08)*

Conclusion

From milestone two, I can conclude that there is a correlation with geographical locations and the rates of stomach and liver cancers. It will require further research of gender differences, in regards to diseases and lifestyles, in the ROK. It will also require research into neighboring regions. *(Appendix 02.06)*

From milestone three, I can conclude that the ROK ranks the highest in hours worked per day; other Asian countries rank above average. The three East-Asian nations that I focused on all have very high RGDPO (real gross domestic product) with China ranked second, Japan ranked third, and the ROK ranked 11th. This is quite high for the ROK considering its dramatically lower population compared to the highest ranking countries.

Assumptions

For milestone two, It was assumed from the start of my analysis of the last dataset, 'death rates among nations', that there could possibly be a regional or even cultural correlation. It just so happens that East Asian nations have the highest rates of stomach and liver cancers. This leads me to believe that regional significance, cultural similarities, and shared histories of different nations have an effect on the subject. I will need to conduct further research to find any significant correlations.

It can also be assumed that the rates of stomach and liver cancers are significantly higher for men than for women. I think that an indication of this would be in the gender inequality of the region. The higher the inequality, the lower the rates of stomach and liver cancers will likely be.

Difference in ethnic and racial attributes could also affect the subject. I think that it could be assumed that alcohol and tobacco have different levels of effects depending on the race of the group. This is something I will explore in the next milestone.

For milestone three, it can be assumed that there is a geographical correlation with stomach cancer in East Asia. I was unable to delve into ethnic and racial considerations. Gender inequality does not seem to correlate with stomach cancer. The nations with the highest gender inequality are mostly African and Middle Eastern nations, which generally do not have exceptionally high levels of stomach cancer.

Limitations

For milestone two, in the last dataset, gender was not included as categorical data. I will need to find additional data to substitute for this. I will also need to find similar data on neighboring countries such as Japan and China.

Several datasets only include statistics on a one to two year timespan. In order to conduct the research required, I will likely need to obtain datasets that cover the entire modern history of the Republic of Korea. This would span from the end of the Korean war until the present day.

In milestone three, since almost all of my datasets contain data for all countries, I was unable to return significant accuracy due to the high variations. This is also likely due to the nature of the data.

I was also unable to research ethnic and racial implications regarding the topic due to time constraint. Based on common knowledge of these implications in other factors, such as alcohol consumption and lactose intolerance, it is likely that this research would be significant for this topic.

Challenges

For milestone two, one major challenge that I am facing is finding one dataset with all the content required for this research. I will likely need to combine content from multiple datasets. I will also need to make sure that correlation is statistically significant in this process. Thus far, I have not

encountered any indication of bias in the data due to the contrast between urban and rural areas; however, this is difficult to detect. Exploring other regions and making comparisons will possibly shed light on this potential issue.

I will also need to consider other metrics to include in my research. Metrics to consider could be the following: gender, race, religion, economy, and standard of living. I will need to obtain sufficient data resources in order to extract the necessary statistical information to conduct further research.

For milestone three, I was unable to return statistically significant results regarding gender inequality. I believe that the dataset is skewed and question how these variables are measured. I think that East Asia has a much higher gender inequality than what the data shows. I would require more time to research further into this.

Future Uses / Additional Applications

Once I have built a model, I plan to use the same approach in exploring other regions based on their cancer rates. I hope to build a model that can predict disease rate trajectories and categorize regions based on a quantitative value of gender equality. Other metrics to consider for this, other than gender could be race, religion, economy, and standard of living.

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 K-Means clustering and PCA.

Ethical Assessment

While working with a large amount of qualitative data, there are many ethical implications. I will need to consider many cultural aspects of each region. This could include surveys. I will need to create models that aren't prone to biases in these instances.

The data, thus far, was not skewed. It's safe to assume that regional differences will possibly become a liability. It was also assumed that due to the stark contrast between urban and rural areas, data may not be available or the data may be incomplete due to faulty record keeping. I will need to delve further into this as I expand my research into other regions. Historical data may also be less accurate in these regions.

I question the results of the gender inequality dataset. I would need to delve further into the ways that these metrics are measured as it is likely a majority of qualitative data. I do think that gender inequality has a universal effect on all facets of a society, and reliable data will be required to research further into this.

References

2021, 2024.06.22, Cancer incident cases and incidence rates by site(24 items), sex, age group https://kosis.kr/statHtml/statHtml.do?orgId=117&tblId=DT_117N_A00023&conn_path=I2&language=en

2022, 2024.06.22, 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

World Health Organization/International Agency for Research on Cancer https://www.kaggle.com/datasets/antimoni/cancer-deaths-by-country-and-type-1990-2016

World Health Organization/International Agency for Research on Cancer https://www.kaggle.com/datasets/ahmadjalalmasood123/worldwide-cancer-data

Appendix

01.01

```
# 02.02.01
# read csv
# dataset 암발생건수및현장별발생률24항목성별연령군_20240614155136.csv
# dt01

dt01_cancer_incidents_____00 = pd.read_csv('암발생건수및현장별발생률24항목성별연령군_20240614155136.csv')
```

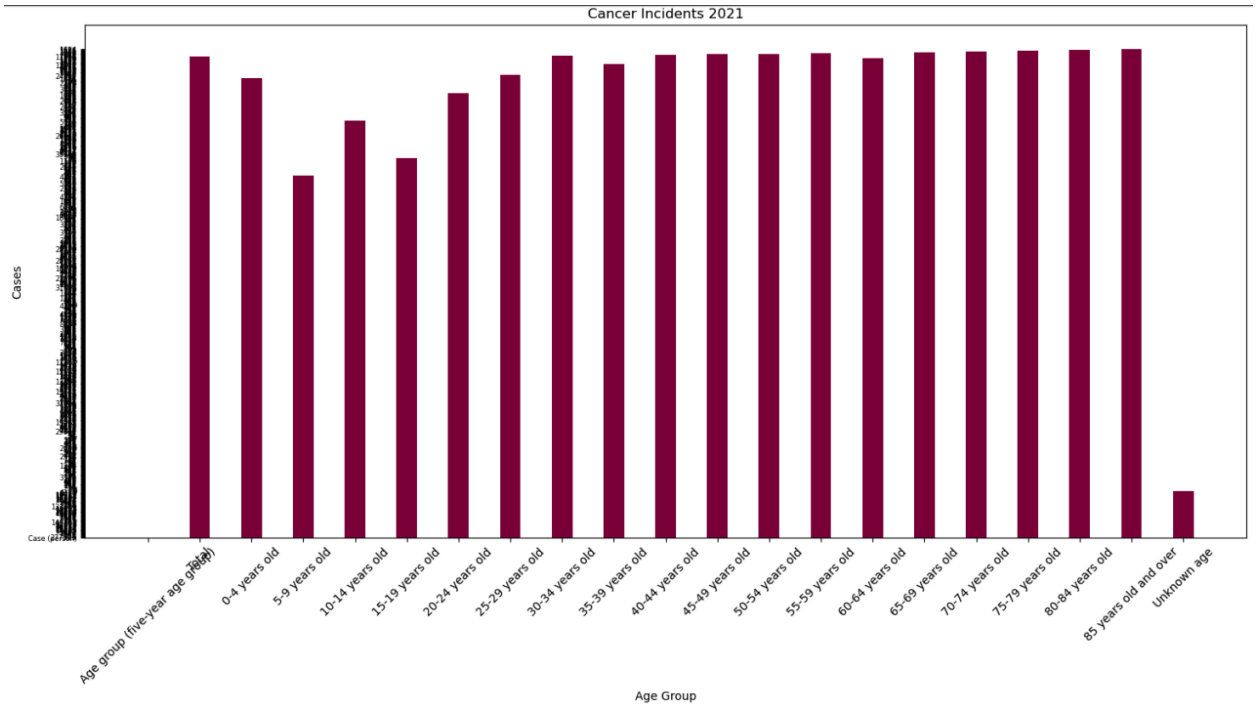
```
# 02.02.01
# read csv
# dataset 원인별사망및사망률104항목성별연령별5년__20240614155018.csv
# dt02

dt02_rates_death_____00 = pd.read_csv('원인별사망및사망률104항목성별연령별5년__20240614155018.csv')
```

```
# 02.02.01
# read csv
# dataset 원인별사망및사망률104항목성별연령별5년__20240614155018.csv
# dt03

dt03_stats_screening_____00 = pd.read_csv('지역별성별별위암검진진단통계_20240614154803.csv')
```

```
# 02.02.03
# return first ten rows
# dt01

dt01_cancer_incidents_____00.head(10)
```

	Cancer Site (24 items)	Sex	Age group (five-year age group)	2021	2021.1
0	Cancer Site (24 items)	Sex	Age group (five-year age group)	Case (person)	Age-Specific Rate (rate per 100000)
1	All Cancers (C00-C96)	Total	Total	277523	540.6

01.02


Cancer Incidents 2021
Cases
Age group (five-year age group)
Total
0-4 years old
5-9 years old
10-14 years old
15-19 years old
20-24 years old
25-29 years old
30-34 years old
35-39 years old
40-44 years old
45-49 years old
50-54 years old
55-59 years old
60-64 years old
65-69 years old
70-74 years old
75-79 years old
80-84 years old
85 years old and over
Unknown age
Age Group

01.03

```
# 02.02.01
# read csv
# dataset Worldwide Cancer Dataset (females) .csv
# dt04

dt04_WW_cancer_DS_fem_______00 = pd.read_csv('Worldwide Cancer Dataset (females) .csv')
```

```
# 02.02.01
# read csv
# dataset Worldwide Cancer Dataset (Males) .csv
# dt05

dt05_WW_cancer_DS_mal_______00 = pd.read_csv('Worldwide Cancer Dataset (Males) .csv')
```

```
# 02.02.01
# read csv
# dataset Worldwide Cancer Dataset.csv
# dt06

dt06_WW_cancer_DS_all_______00 = pd.read_csv('Worldwide Cancer Dataset.csv')
```

```
# 02.02.03
# return first ten rows
# dt04

dt04_WW_cancer_DS_fem_______00.head(10)
```

	Rank	Cancer	New cases in 2020	% of all cancers
0	NaN	All cancers*	8,751,759	NaN

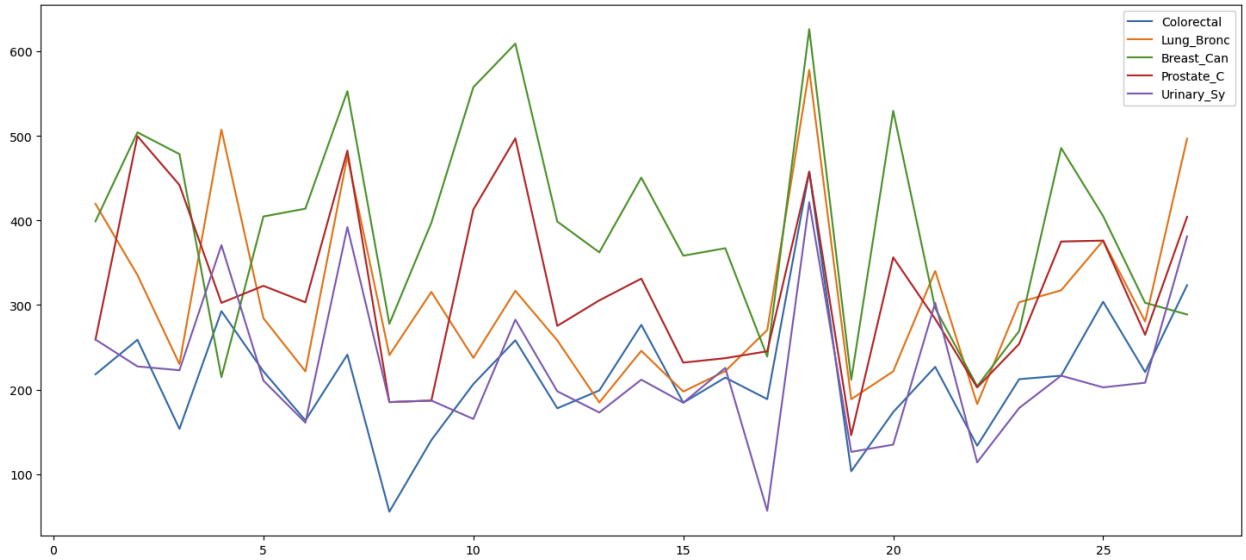
01.04

```
# 02.02.01
# read csv
# dataset Worldwide Cancer Dataset.csv
# dt07

dt07_rates_cancer__________00 = pd.read_csv('Cancer_Rates.csv')
```

```
# 02.02.03
# return first ten rows
# dt07

dt07_rates_cancer__________00.head(10)
```

	FID	ZIP	Colorectal	Lung_Bronc	Breast_Can	Prostate_C	Urinary_Sy	All_Cancer	SHAPE_Length	SHAPE_Area
0	1	60002	218.062127	419.666735	399.094836	259.205925	259.205925	2703.147501	215525.155184	1.149062e+09

01.05



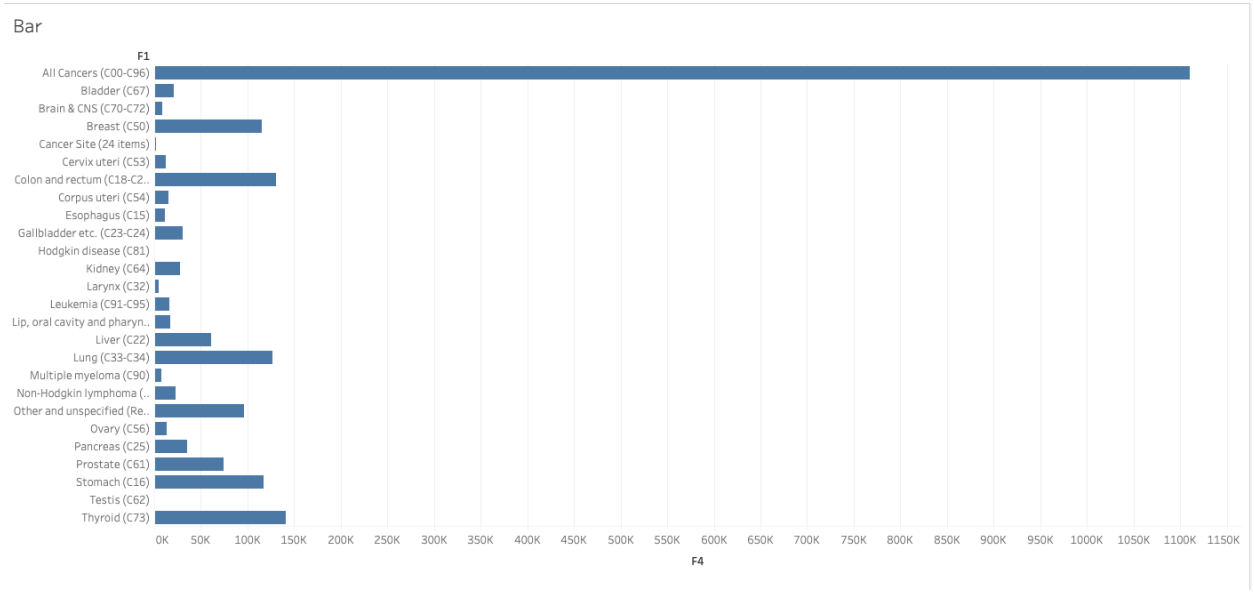
01.06

```
# 02.02.01
# read csv
# dataset Worldwide Cancer Dataset.csv
# dt08

dt08_deaths_cancer_country__00 = pd.read_csv('Cancer Deaths by Country and Type Dataset.csv')
```

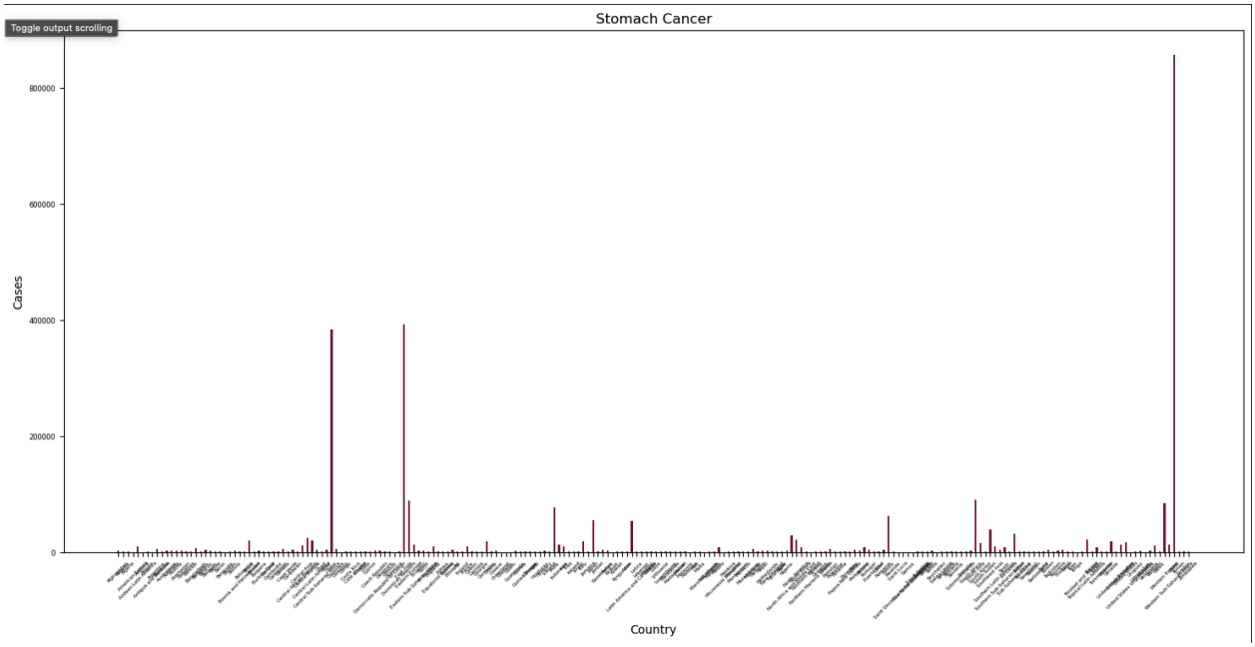
```
# 02.02.03
# return first and last ten rows
# dt08

print(dt08_deaths_cancer_country__00.head(10))
print(dt08_deaths_cancer_country__00.tail(10))
```

```
       Country Code  Year  ...  Hodgkin lymphoma   Multiple myeloma   Other cancers
0  Afghanistan  AFG  1990  ...         191.367386          50.719442      294.839679
1  Afghanistan  AFG  1991  ...         203.509622          54.317640      311.469065
```

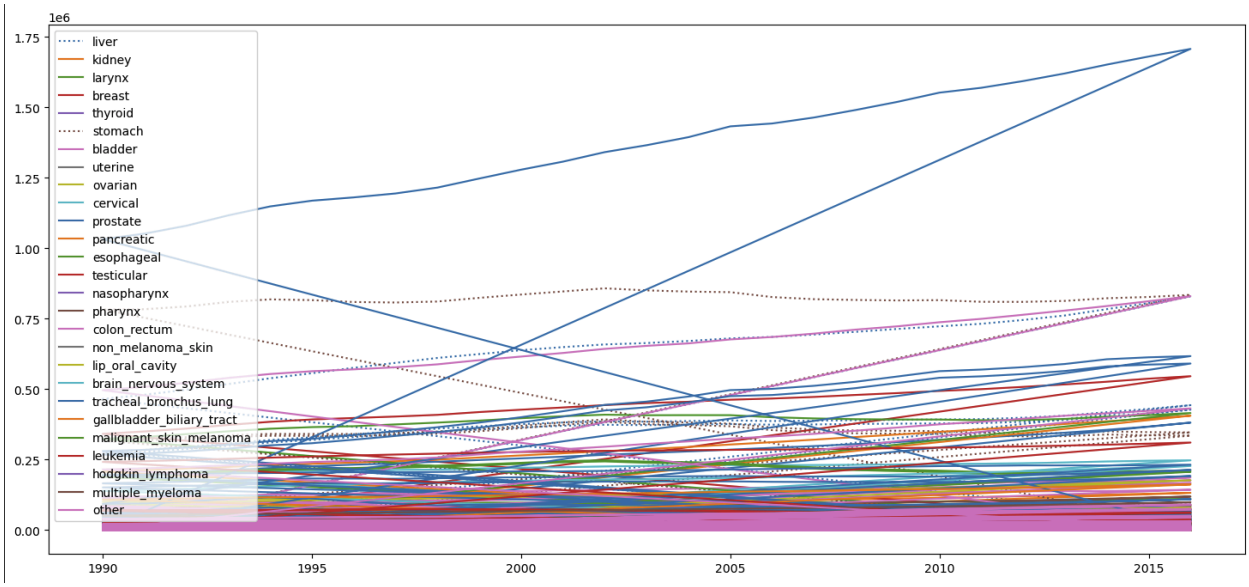
02.01


Bar
F1
All Cancers (C00-C96)
Bladder (C67)
Brain & CNS (C70-C72)
Breast (C50)
Cancer Site (24 items)
Cervix uteri (C53)
Colon and rectum (C18-C2..
Corpus uteri (C54)
Esophagus (C15)
Gallbladder etc. (C23-C24)
Hodgkin disease (C81)
Kidney (C64)
Larynx (C32)
Leukemia (C91-C95)
Lip, oral cavity and pharyn..
Liver (C22)
Lung (C33-C34)
Multiple myeloma (C90)
Non-Hodgkin lymphoma (..
Other and unspecified (Re..
Ovary (C56)
Pancreas (C25)
Prostate (C61)
Stomach (C16)
Testis (C62)
Thyroid (C73)
0K 50K 100K 150K 200K 250K 300K 350K 400K 450K 500K 550K 600K 650K 700K 750K 800K 850K 900K 950K 1000K 1050K 1100K 1150K
F4

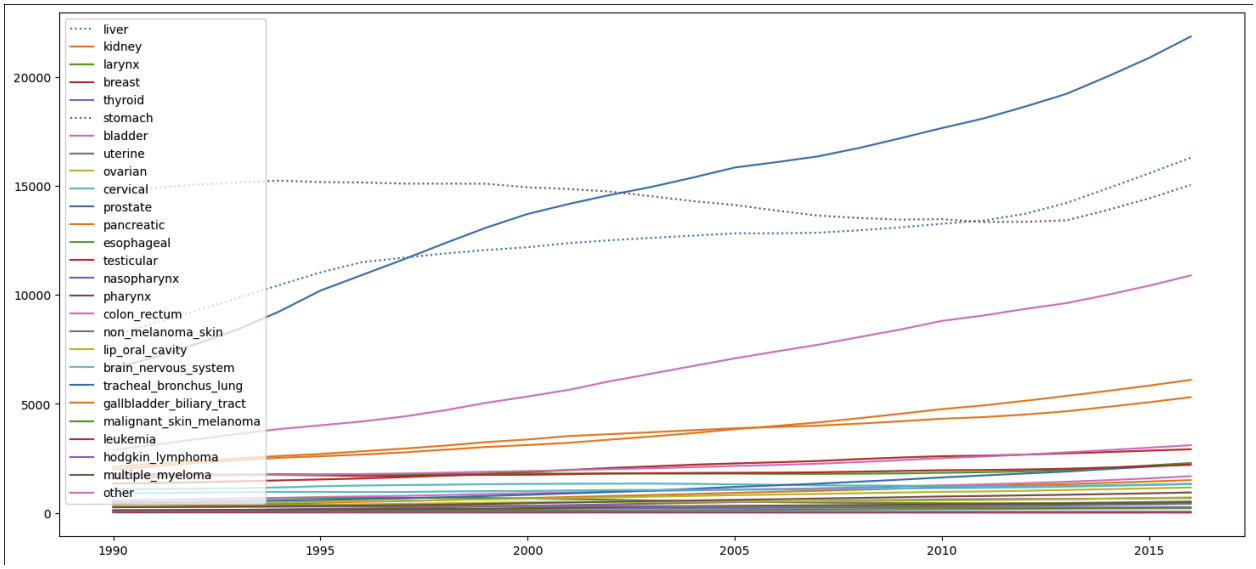
02.02


Toggle output scrolling
Stomach Cancer
800000
600000
400000
200000
0
Cases
Country

02.03

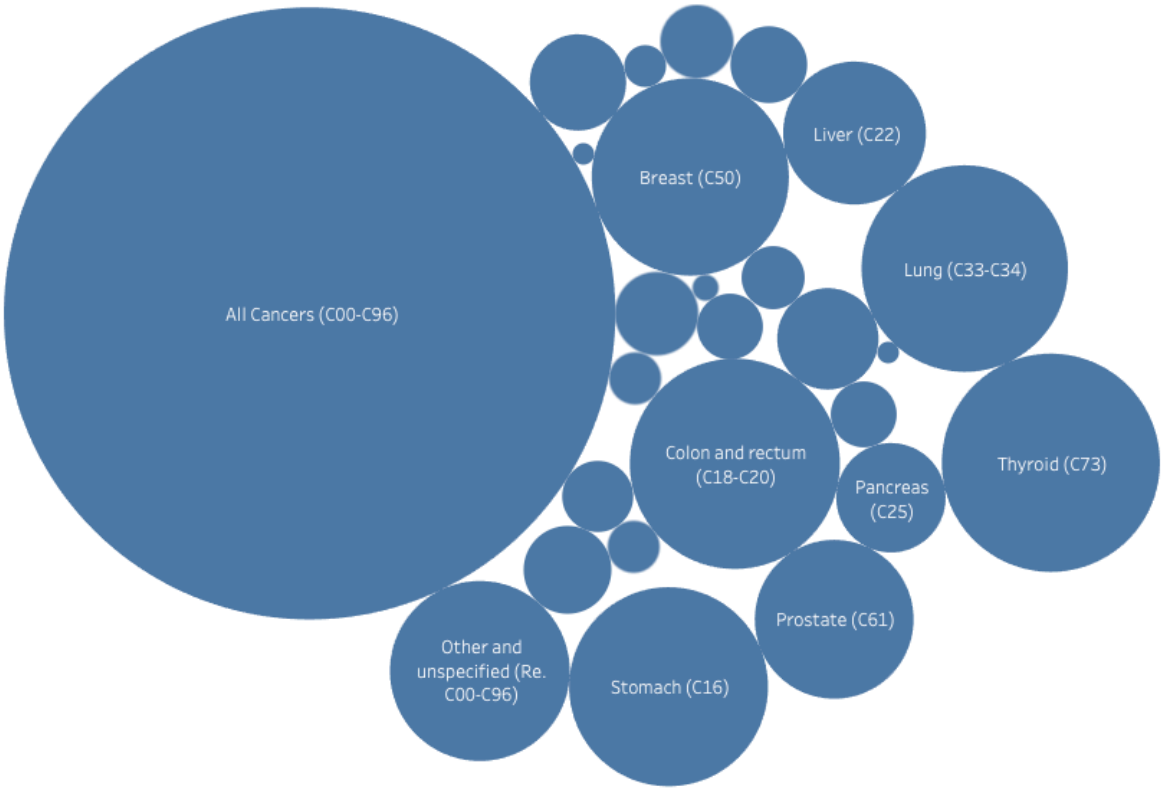

1e6
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
1990
1995
2000
2005
2010
2015
liver
kidney
larynx
breast
thyroid
stomach
bladder
uterine
ovarian
cervical
prostate
pancreatic
esophageal
testicular
nasopharynx
pharynx
colon_rectum
non_melanoma_skin
lip_oral_cavity
brain_nervous_system
tracheal_bronchus_lung
gallbladder_biliary_tract
malignant_skin_melanoma
leukemia
hodgkin_lymphoma
multiple_myeloma
other

02.04

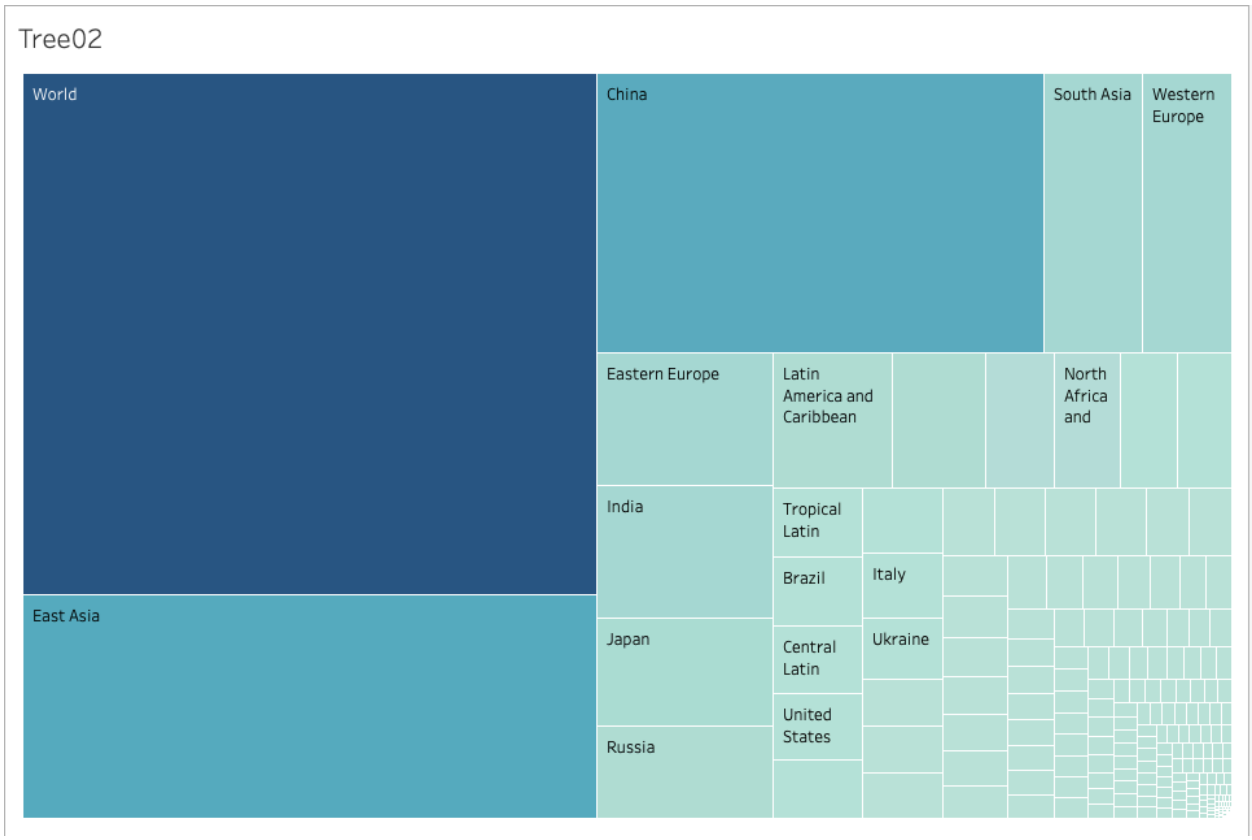

20000
15000
10000
5000
0
1990
1995
2000
2005
2010
2015
liver
kidney
larynx
breast
thyroid
stomach
bladder
uterine
ovarian
cervical
prostate
pancreatic
esophageal
testicular
nasopharynx
pharynx
colon_rectum
non_melanoma_skin
lip_oral_cavity
brain_nervous_system
tracheal_bronchus_lung
gallbladder_biliary_tract
malignant_skin_melanoma
leukemia
hodgkin_lymphoma
multiple_myeloma
other

02.05

Bubble

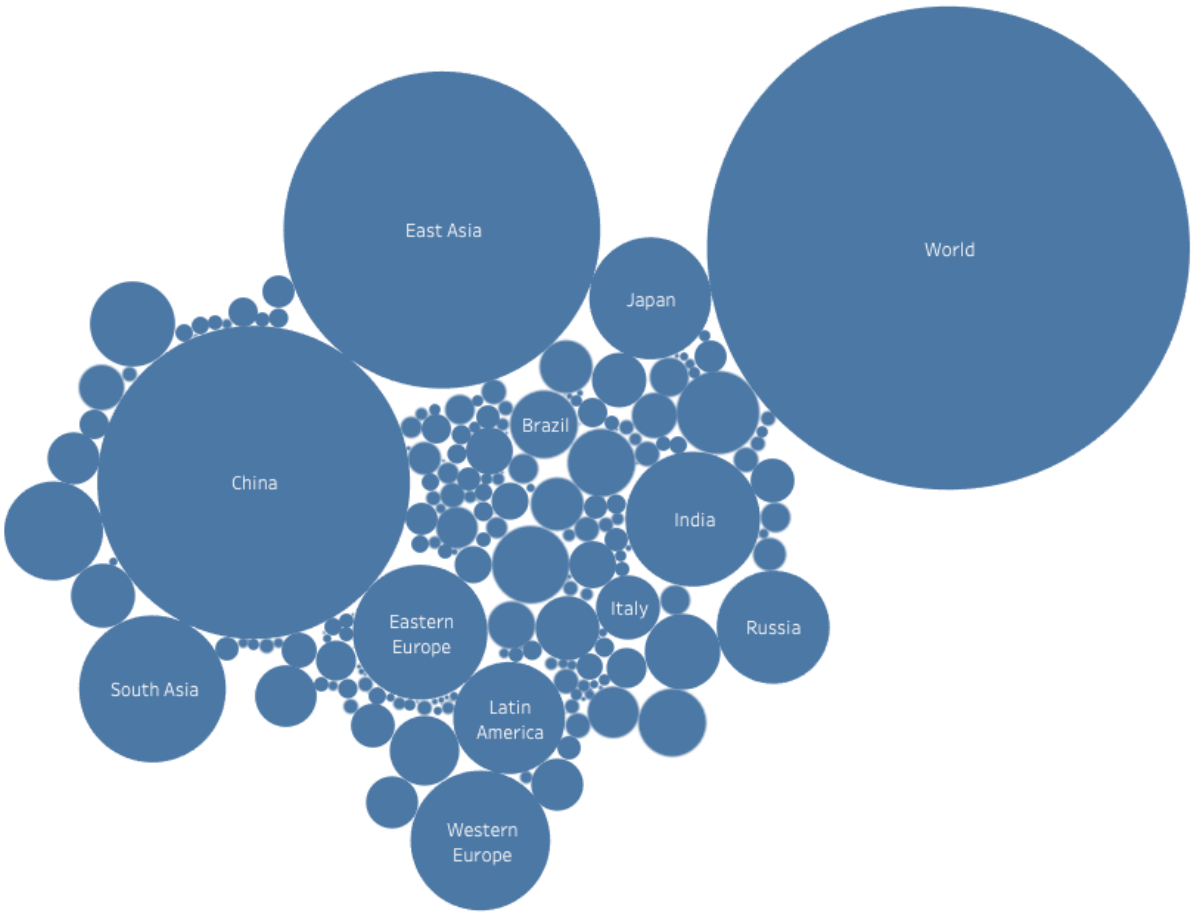

All Cancers (C00-C96)
Breast (C50)
Liver (C22)
Lung (C33-C34)
Colon and rectum (C18-C20)
Pancreas (C25)
Thyroid (C73)
Prostate (C61)
Other and unspecified (Re. C00-C96)
Stomach (C16)

02.06

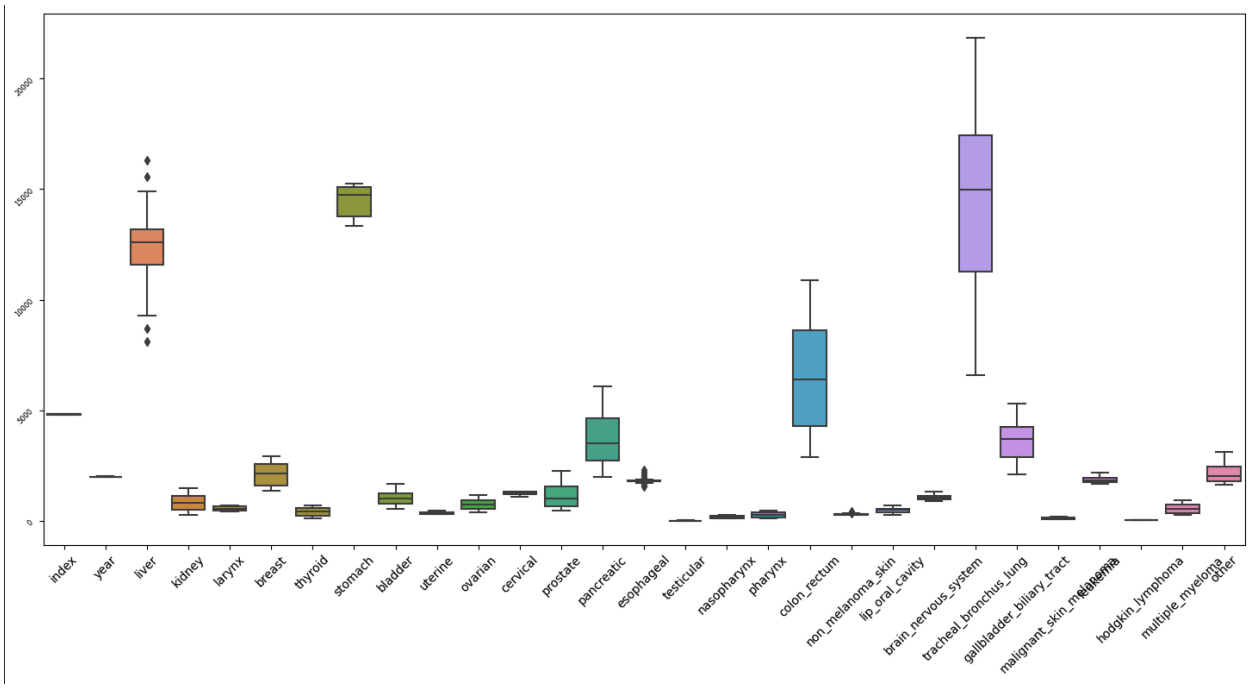

Tree02
World
East Asia
China
South Asia
Western Europe
Eastern Europe
Latin America and Caribbean
North Africa and
India
Tropical Latin
Brazil
Italy
Japan
Central Latin
Ukraine
United States
Russia

02.07

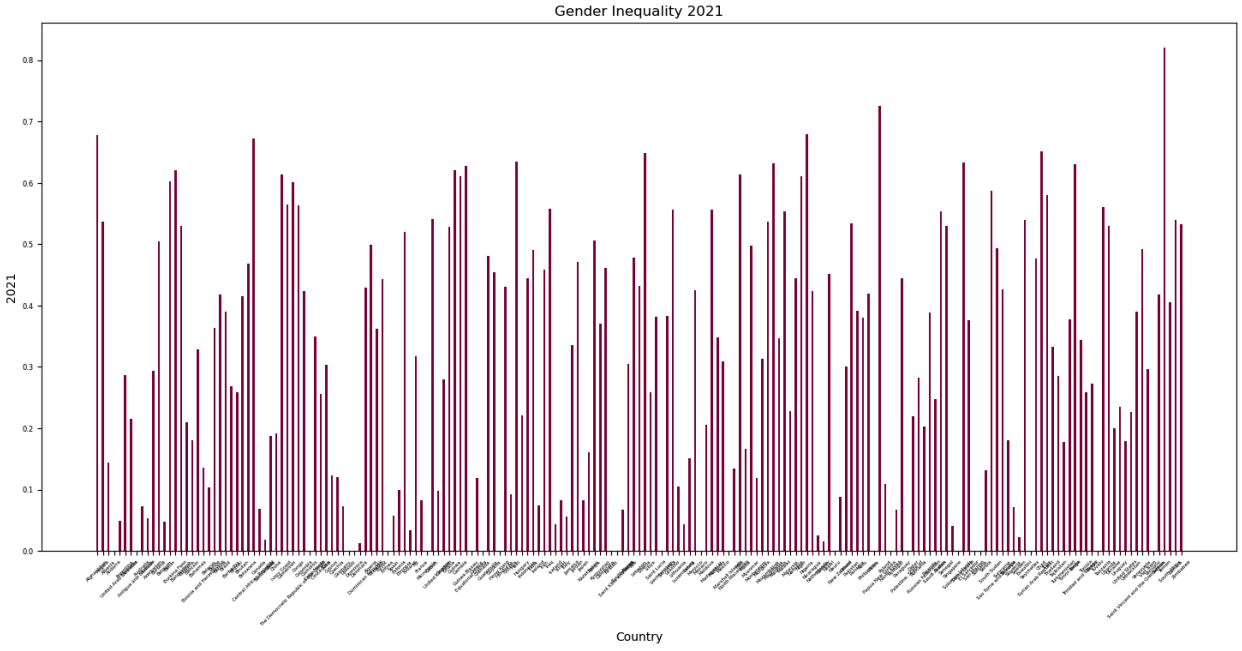
Tree02


East Asia
World
Japan
Brazil
China
India
Italy
Russia
Eastern Europe
South Asia
Latin America
Western Europe

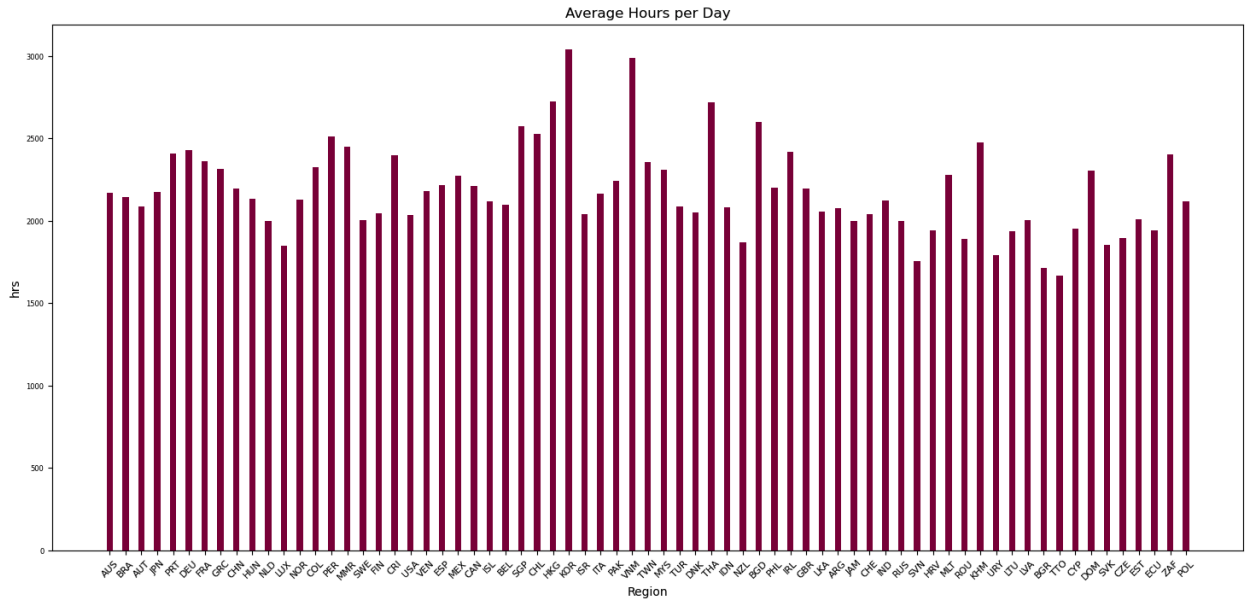
03.01


20000
15000
10000
5000
0
index
year
liver
kidney
larynx
breast
thyroid
stomach
bladder
uterine
ovarian
cervical
prostate
pancreatic
esophageal
testicular
nasopharynx
pharynx
colon_rectum
non_melanoma_skin
lip_oral_cavity
brain_nervous_system
tracheal_bronchus_lung
gallbladder_biliary_tract
malignant_skin_melanoma
hodgkin_lymphoma
multiple_myeloma
other

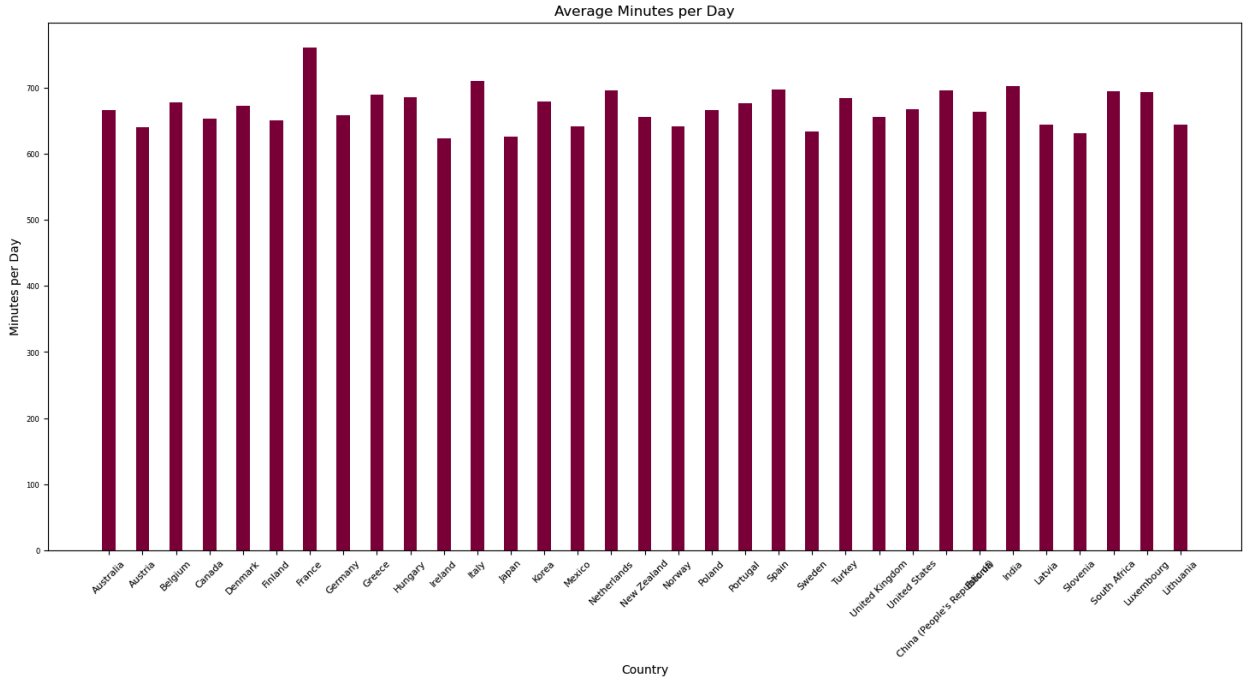
03.02


Gender Inequality 2021
2021
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Country

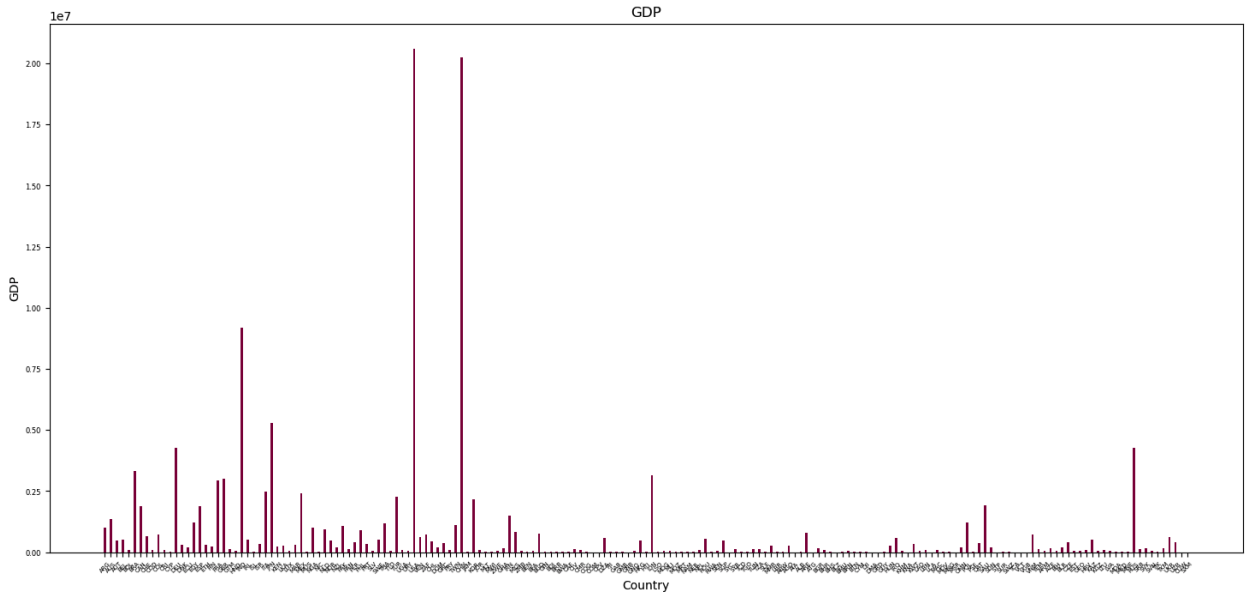
03.03


Average Hours per Day
hrs
Region
AUS BRA AUT JPN PRT DEU FRA GRC CHN HUN NLD LUX NOR COL PER MMR SWE FIN CRI USA VEN ESP MEX CAN ISL BEL SGP CHL HKG KOR ISR ITA PAK VNM TWN MYS TUR DNK THA IDN NZL BGD PHL IRL GBR UKA ARG JAM CHE IND RUS SVN HRV MLT ROU KHM URY LTU LVA BGR TTO CYP DOM SVK CZE EST ECU ZAF POL

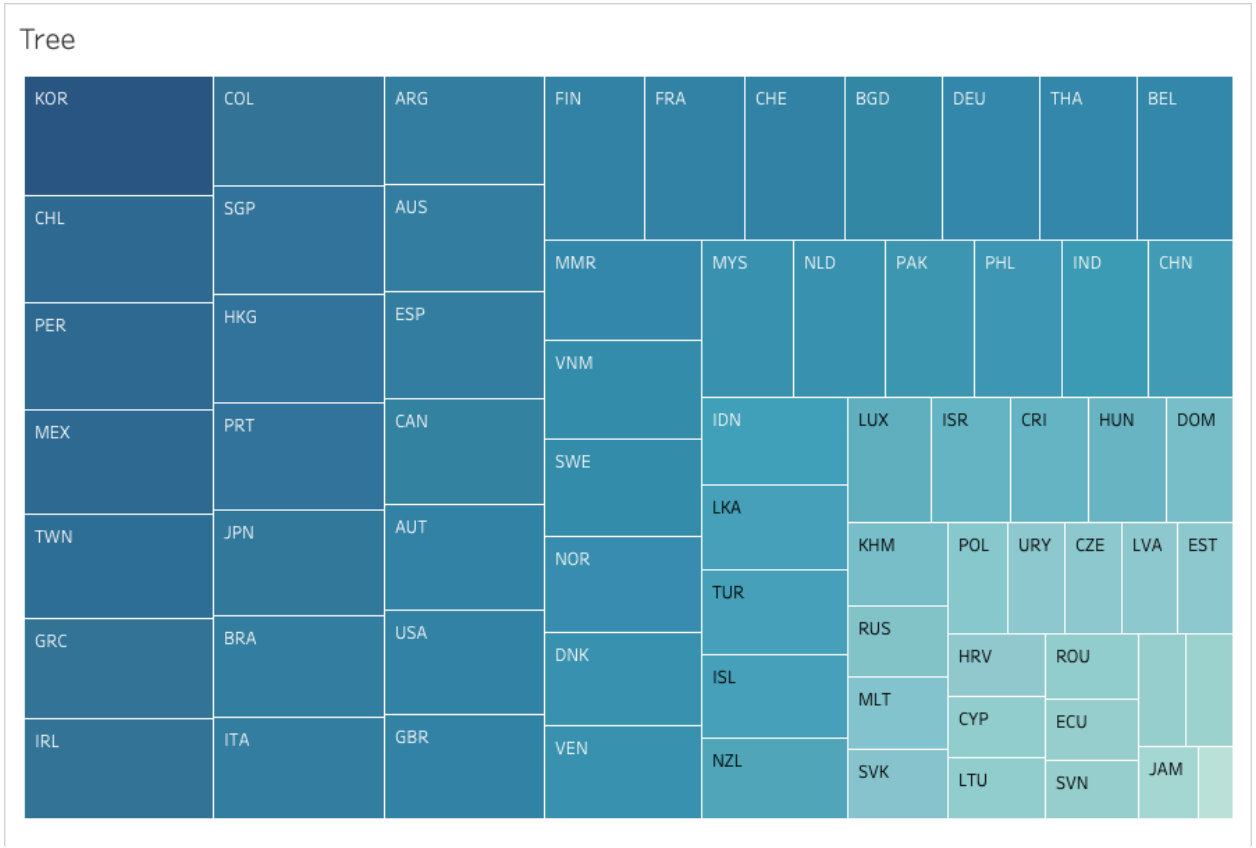
03.04


Average Minutes per Day
Minutes per Day
Country
Australia Austria Belgium Canada Denmark Finland France Germany Greece Hungary Ireland Italy Japan Korea Mexico Netherlands New Zealand Norway Poland Portugal Spain Sweden Turkey United Kingdom United States India Latvia Slovenia South Africa Luxembourg Lithuania

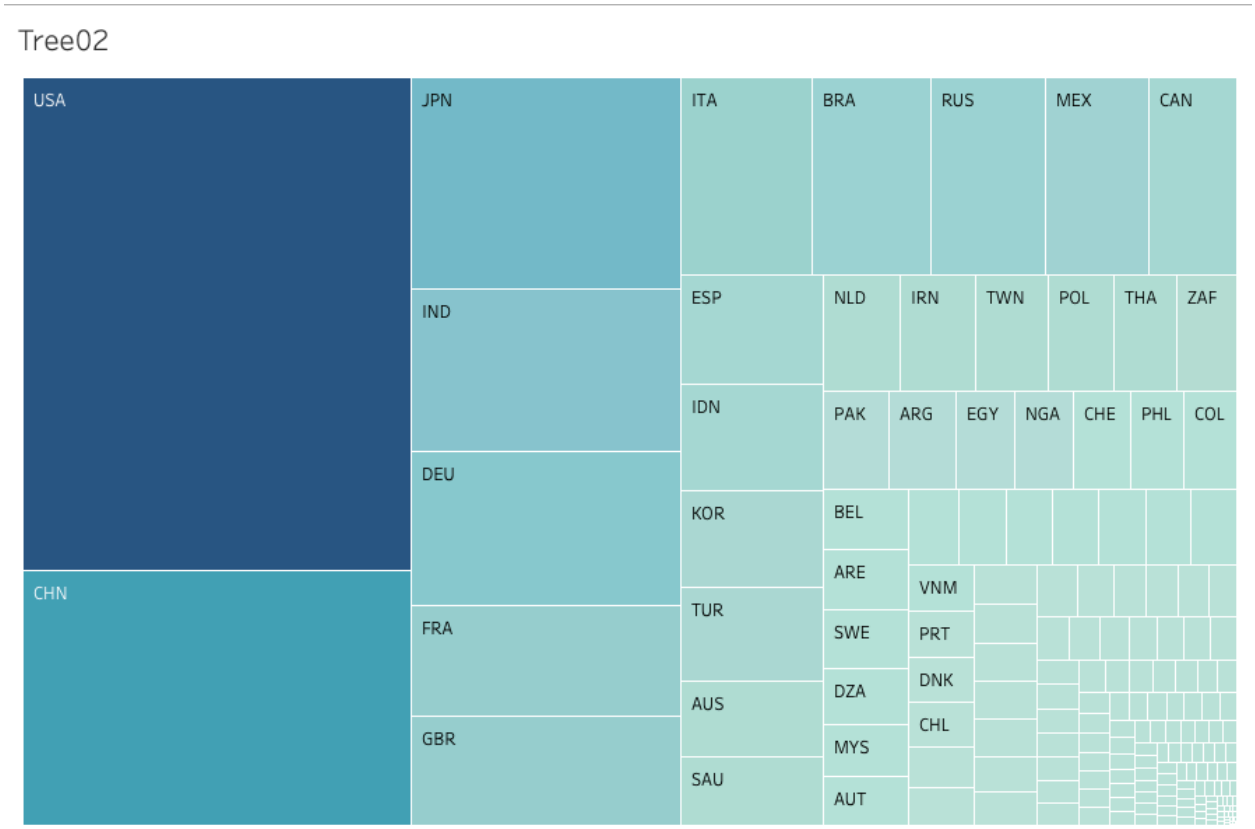
03.05


GDP
1e7
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
GDP
Country

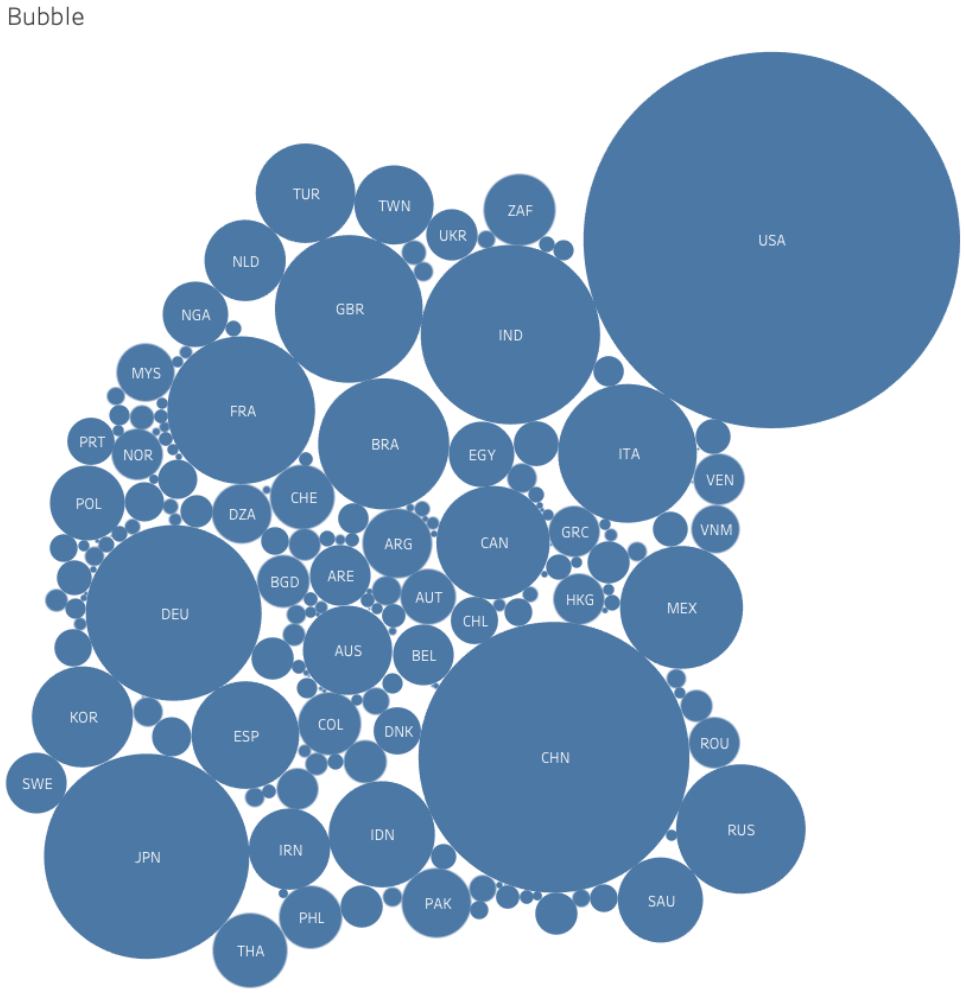
03.06


Tree
KOR
COL
ARG
FIN
FRA
CHE
BGD
DEU
THA
BEL
CHL
SGP
AUS
MMR
MYS
NLD
PAK
PHL
IND
CHN
PER
HKG
ESP
VNM
MEX
PRT
CAN
IDN
LUX
ISR
CRI
HUN
DOM
SWE
LKA
TWN
JPN
AUT
NOR
KHM
POL
URY
CZE
LVA
EST
TUR
GRC
BRA
USA
DNK
RUS
HRV
ROU
ISL
MLT
CYP
ECU
IRL
ITA
GBR
VEN
NZL
SVK
LTU
SVN
JAM

03.07


Tree02
USA
JPN
ITA
BRA
RUS
MEX
CAN
IND
ESP
NLD
IRN
TWN
POL
THA
ZAF
IDN
PAK
ARG
EGY
NGA
CHE
PHL
COL
DEU
KOR
BEL
CHN
ARE
VNM
TUR
SWE
PRT
FRA
DNK
DZA
AUS
CHL
GBR
MYS
SAU
AUT

03.08


Bubble
TUR
TWN
ZAF
UKR
USA
NLD
GBR
IND
NGA
MYS
FRA
BRA
EGY
ITA
PRT
NOR
VEN
POL
DZA
CHE
ARG
CAN
GRC
VNM
BGD
ARE
AUT
HKG
MEX
DEU
CHL
AUS
BEL
KOR
ESP
COL
DNK
CHN
ROU
SWE
RUS
JPN
IRN
IDN
PAK
SAU
PHL
THA

03.09

```
# 04.01.01
# read csv
# dataset Gender Inequality Index.csv
# dt09

dt09_index_gend_ineq________00 = pd.read_csv('Gender Inequality Index.csv')
```

```
# 04.01.02
# read csv
# dataset Avg_hours_worked_(1950-2017).csv
# dt10

dt10_ave_hrs_worked_________00 = pd.read_csv('Avg_hours_worked_(1950-2017).csv')
```

```
# 04.01.03
# read csv
# dataset Time_use_OECD.csv
# dt11

dt11_time_use_oecd__________00 = pd.read_csv('Time_use_OECD.csv')
```

```
# 04.01.04
# read csv
# dataset Yearly_RGDPO_(1950-2017).csv
# dt12

dt12_yearly_rgdpo___________00 = pd.read_csv('Yearly_RGDPO_(1950-2017).csv')
```

Dataset 01: ‘암발 생건수 및 현장 별발생률 24항목 성별 연령군_20240614155136.csv’ as ‘dt01_cancer_incidents_____00’

Dataset 02: ‘원인별사망 및 사망률 104 항목 성별 연령별 5년__20240614155018.csv’ as ‘dt02_rates_death_____00’

Dataset 03: '지역 별 성별 별위 암검진 진단 통계_____20240614154803.csv' as ‘dt03_stats_screening_____00’

Dataset 04: 'Worldwide Cancer Dataset (females) .csv' as ‘dt04_WW_cancer_DS_fem_____00’

Dataset 05: 'Worldwide Cancer Dataset (Males) .csv' as ‘dt05_WW_cancer_DS_mal_____00’

Dataset 06: 'Worldwide Cancer Dataset.csv' as ‘dt06_WW_cancer_DS_all_____00’

Dataset 07: 'Cancer_Rates.csv' as ‘dt07_rates_cancer_____00’

Dataset 08: 'Cancer Deaths by Country and Type Dataset.csv' as ‘dt08_deaths_cancer_country__00’

Dataset 09: 'Gender Inequality Index.csv' as ‘dt09_index_gend_ineq_____00’

Dataset 10: 'Avg_hours_worked_(1950-2017).csv' as ‘dt10_ave_hrs_worked_____00’

Dataset 11: 'Time_use_OECD.csv' as ‘dt11_time_use_oecd_____00’

Dataset 12: 'Yearly_RGDPO_(1950-2017).csv' as ‘dt12_yearly_rgdpo_____00’

Ten questions an audience would ask you:

1. How many regions do you think you should include in your research?
    a. I initially focused on three regions as they are geographically relevant to each other. The region is East Asian, however, other regions whose countries have similarities could also be considered.
2. What cultural or historical aspects affect your decisions on the number of inputs?
    a. I wasn't able to conduct research into cultural and historical considerations. If I were to conduct further research, I would especially consider the cultural similarities between the ROK and Japan, and the ROK and China. The ROK and Japan have somewhat shared histories, and the ROK and China have cultural links with one another.
3. What different cultural considerations should be included in your research?
    a. Religion, history, and regional customs should all be considered in this research. In this case, religious similarities between the ROK and China would be relevant while the shared histories of the ROK and Japan should also be relevant.

4. How does the rate at which regions have industrialized affect your research?
    a. The regions that I focused on all industrialized rapidly in modern history. I think this does have a significant impact on this research. For example, nations that have recently industrialized, tend to have a higher rate of smokers among their populations.
5. How significantly does gender inequality affect your research?
    a. Based on the data, gender inequality does not have the assumed effect on this research. I do, however, think that the dataset is inaccurate. I would need more time to investigate this. I think that the dataset I obtained is likely based on qualitative data which spawns ethical concerns.
6. How do disparities in gender inequality differ among the regions you are researching?
    a. Gender inequality tended to be higher in African and Middle Eastern countries. Counter intuitively, East Asian countries ranked lower, nearly as low as western nations. I think that this data is likely skewed and inaccurate.

7. What are the major indicators of gender inequality based on your research?
    a. The data did not include the metrics used for these calculations. I would need to research further into how the data was obtained and transformed. I think this dataset is skewed due to qualitative data.
8. How much has religion in the past affected the subject you are researching?
    a. I was unable to research religious implications on the subject. I do think that religion has a significant impact on the topic. It definitely deserved further investigation.
9. How has politics affected the subject you are researching?
    a. I was unable to research political implications on the subject. Though likely not as significant as the religious aspect, I do think that politics have a significant impact on the topic, and like religion, should be investigated.

10. After deep insights, what other regions would you consider researching in the future?

    a. I would consider researching any region with an above-average rate of stomach cancer. East Asia is among the highest. There are other regions such as South Asia and Eastern Europe that also have high rates. I would also consider researching other regions with a high rate of other cancers; for example, lung cancer seems to rank higher in developing nations. There must be a correlation with development and smoking. This would be worth researching.