DSC680 - Applied Data Science

24.06.30 Week Four

Milestone Three

Professor Iranitalab

Ross Kim-Schreck

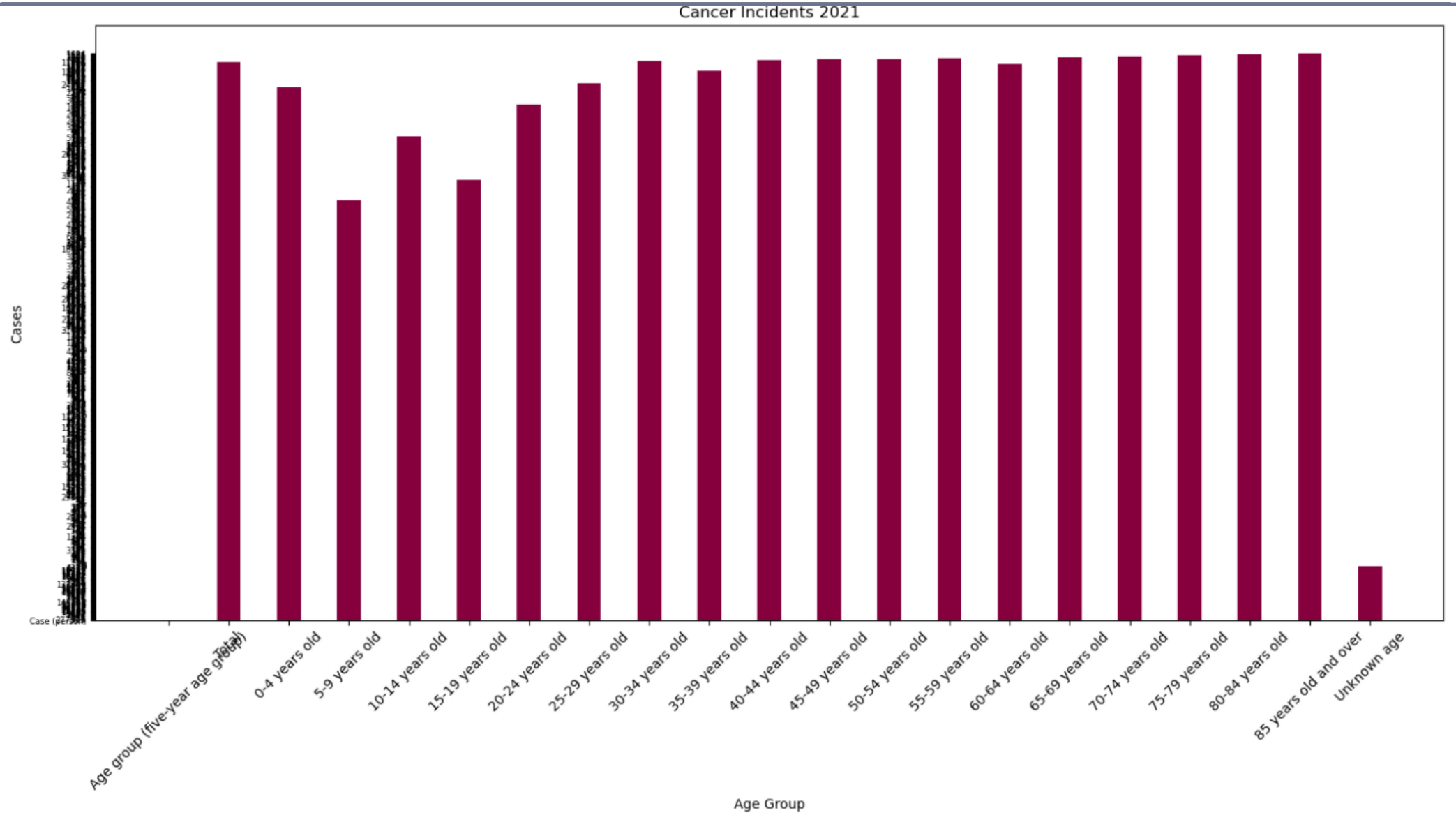
```python
# 02.02.01
# read csv
# dataset 암발생건수및현장별발생률24항목성별연령군_20240614155136.csv
# dt01

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

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

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

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

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

```python
# 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


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

```
# 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')
```

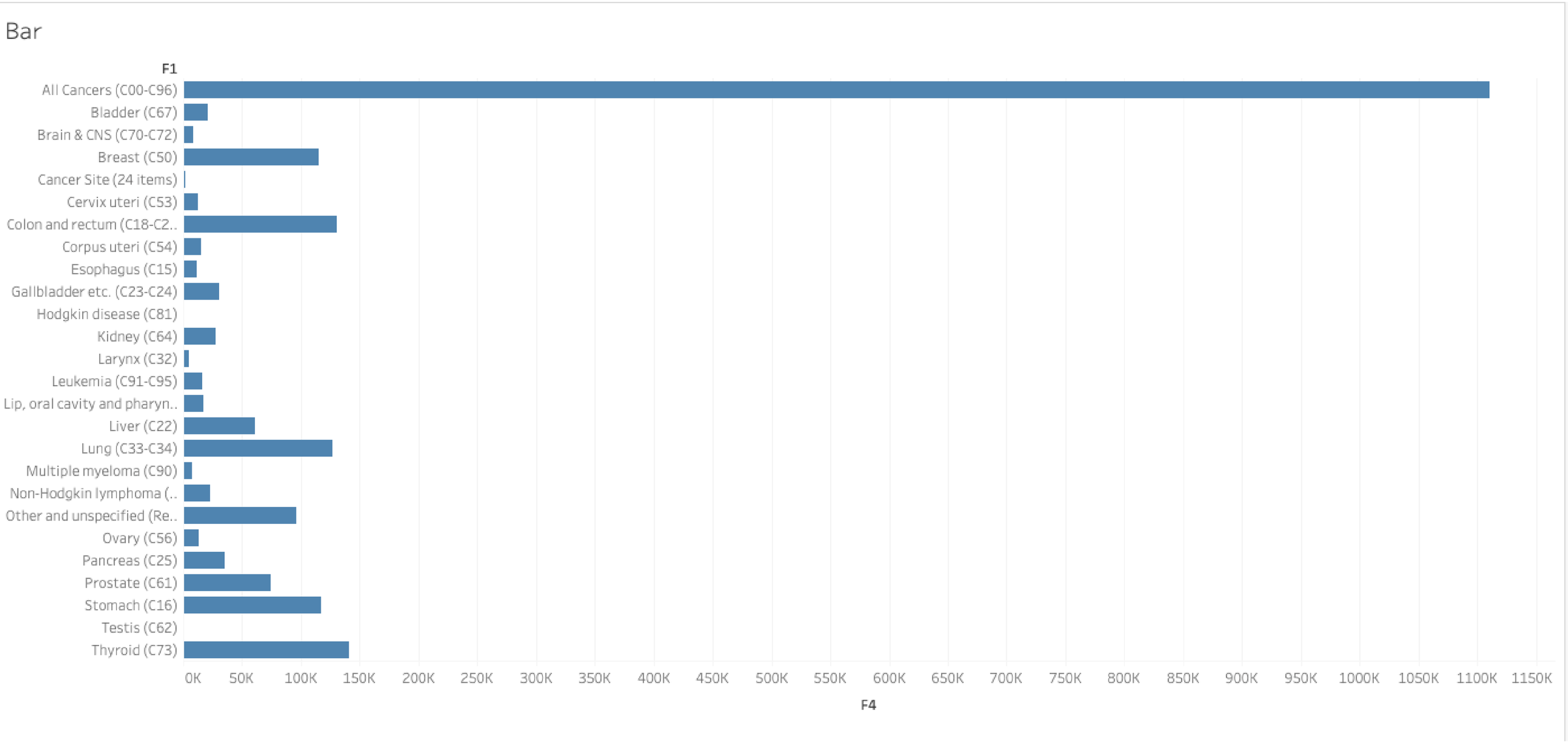
```
# 02.02.03
# return first ten rows
# dt04

dt04_WW_cancer_DS_fem_______00.head(10)
```

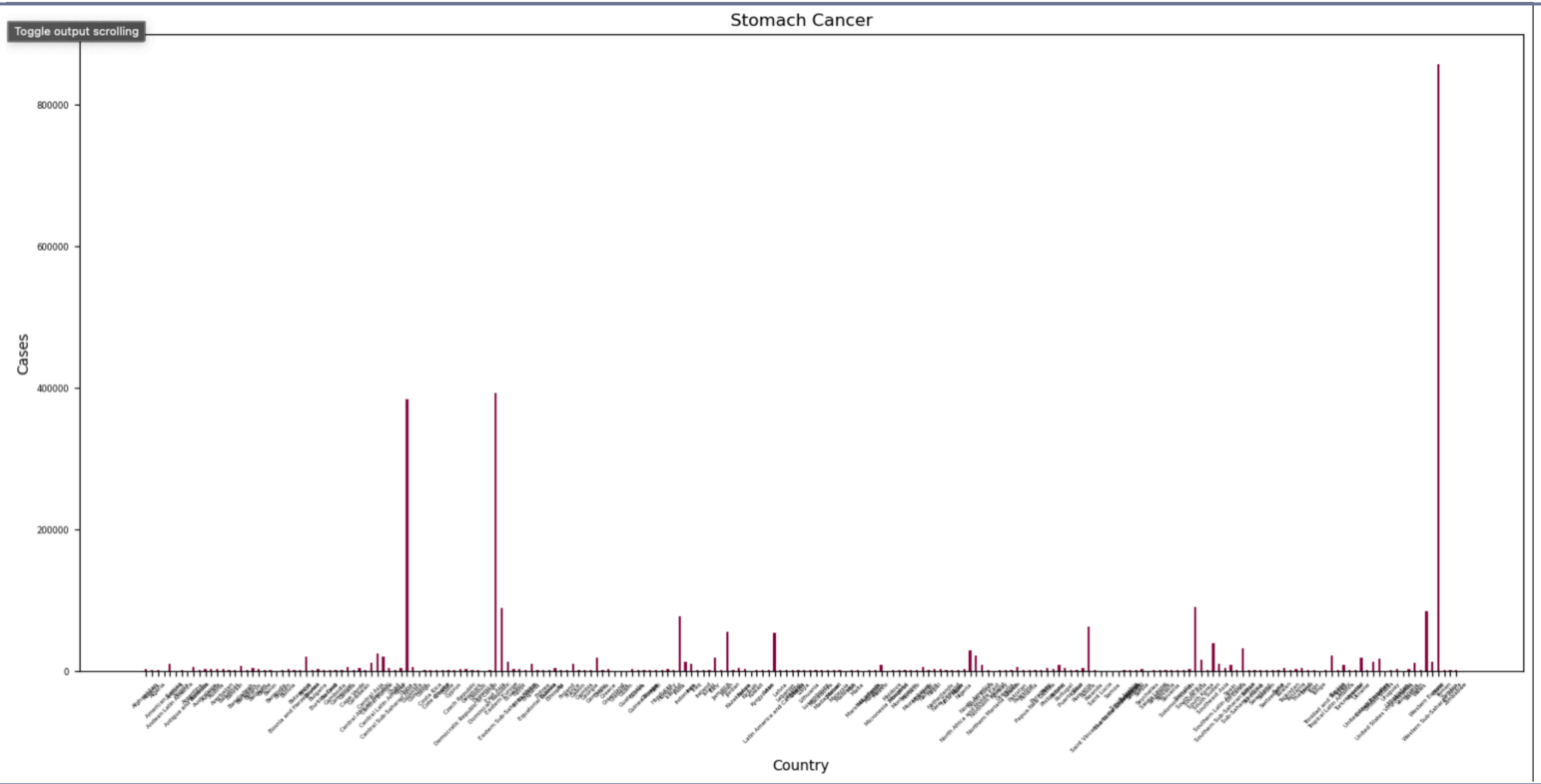
	Rank	Cancer	New cases in 2020	% of all cancers
0	NaN	All cancers*	8,751,759	NaN

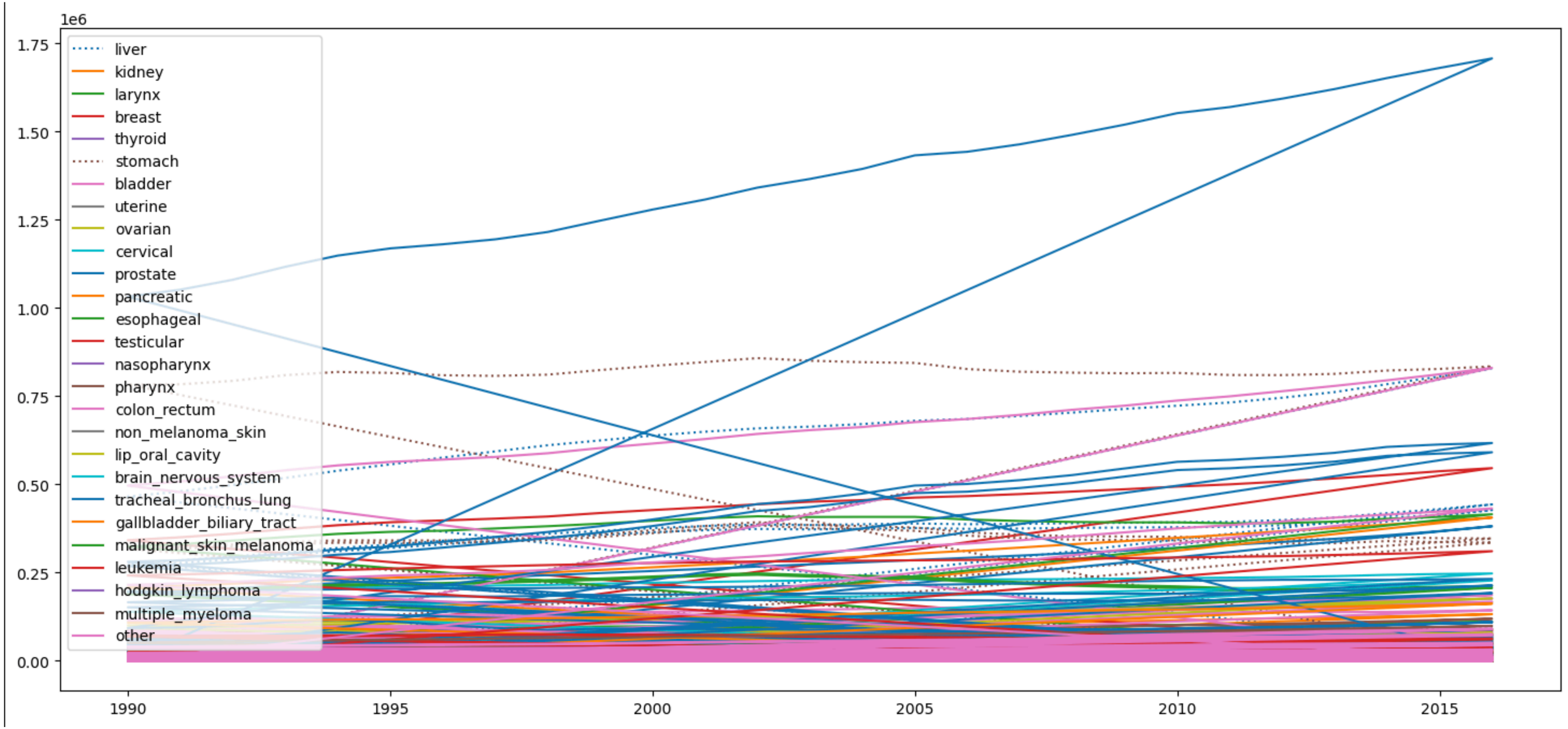
```
# 02.02.01
# read csv
# dataset Worldwide Cancer Dataset.csv
# dt07

dt07_rates_cancer___________00 = pd.read_csv('Cancer_Rates.csv')
```

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

dt07_rates_cancer___________00.head(10)
```

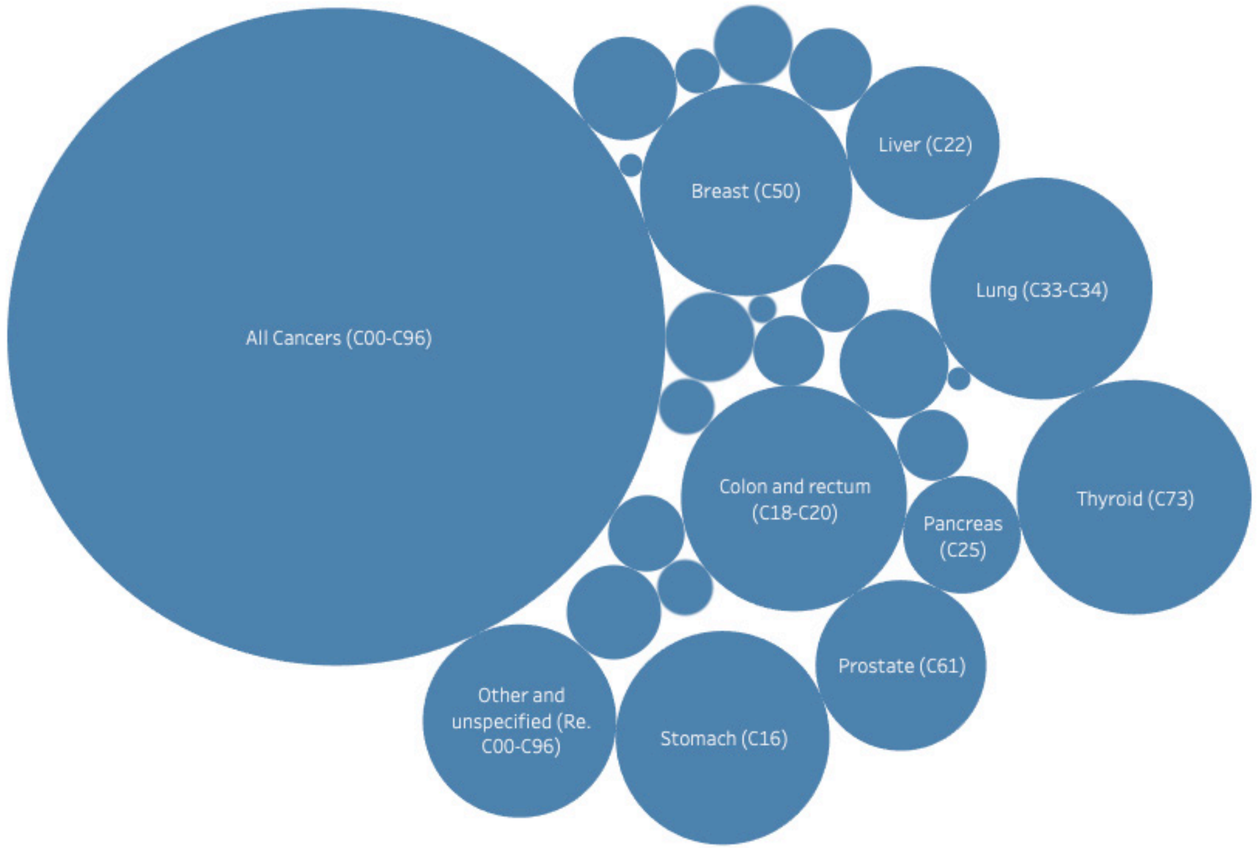
	FID	ZIP	Colorectal	Lung_Bronc	Breast_Can	Prostate_C	Urinary_Sy	All_Cancer	SHAPE_Length	SHAPE_Area
0	1	60002	218.062127	419.666735	399.094836	259.205925	259.205925	2703.147501	215525.155184	1.149062e+09


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

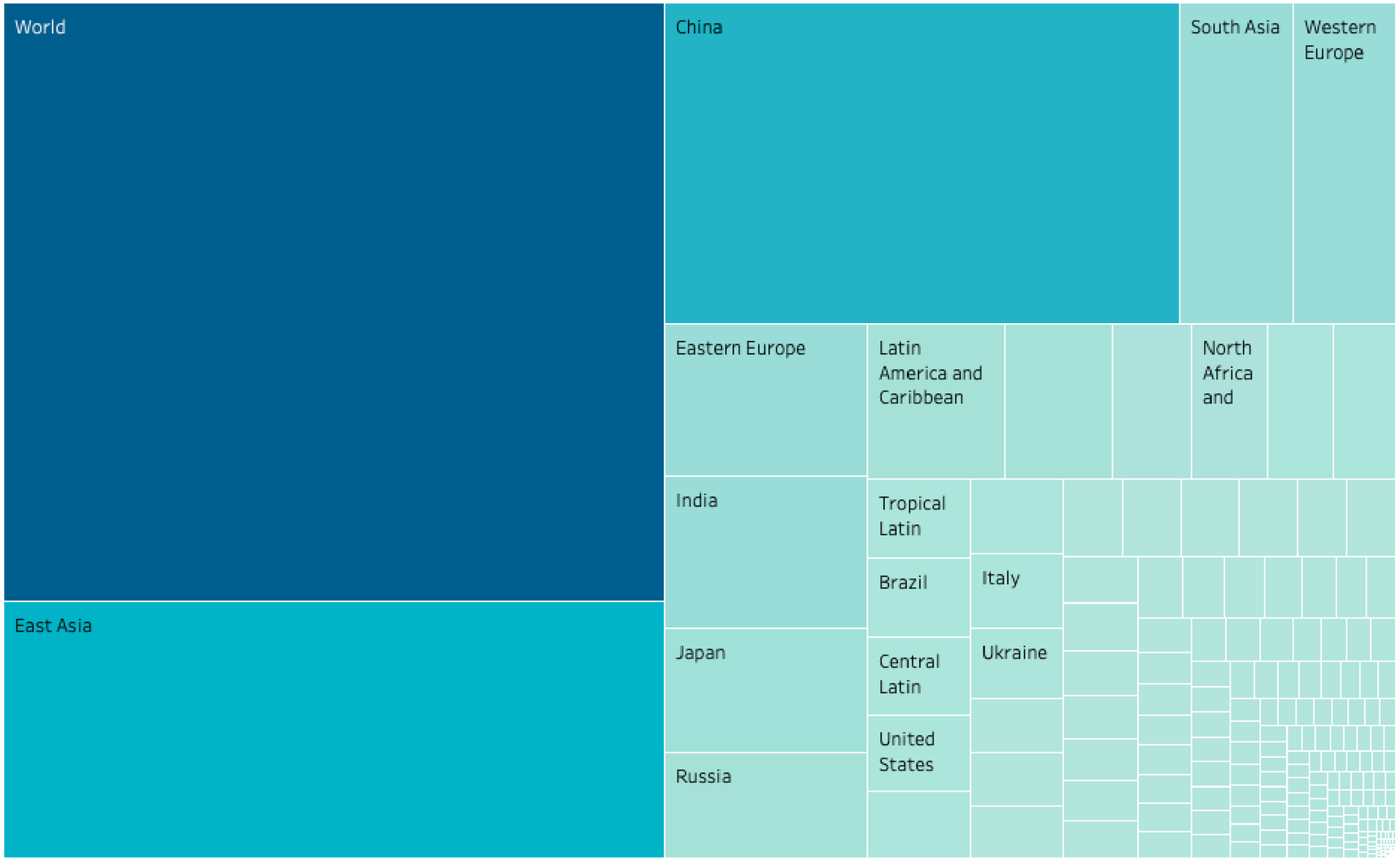

Toggle output scrolling
Stomach Cancer
Cases
800000
600000
400000
200000
0
Country


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

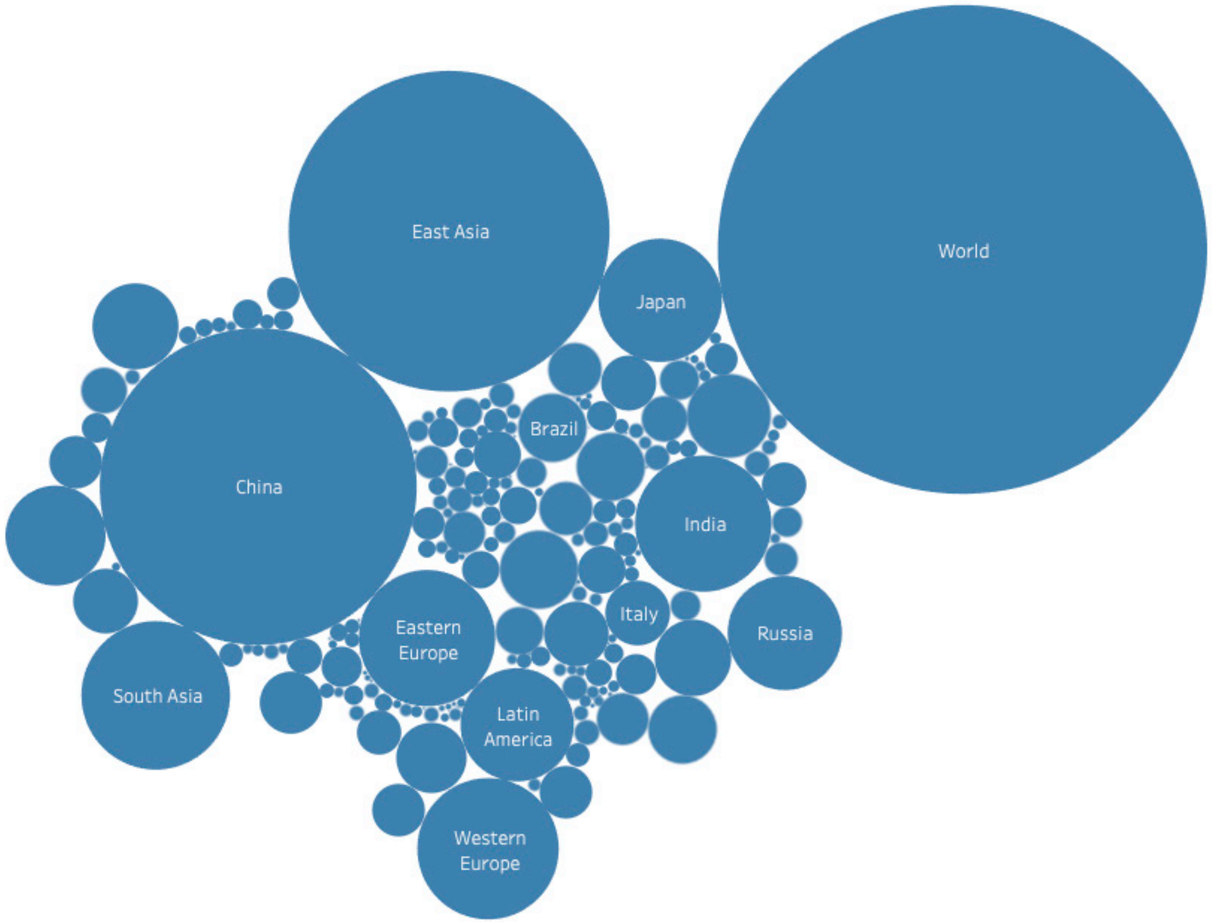
Bubble

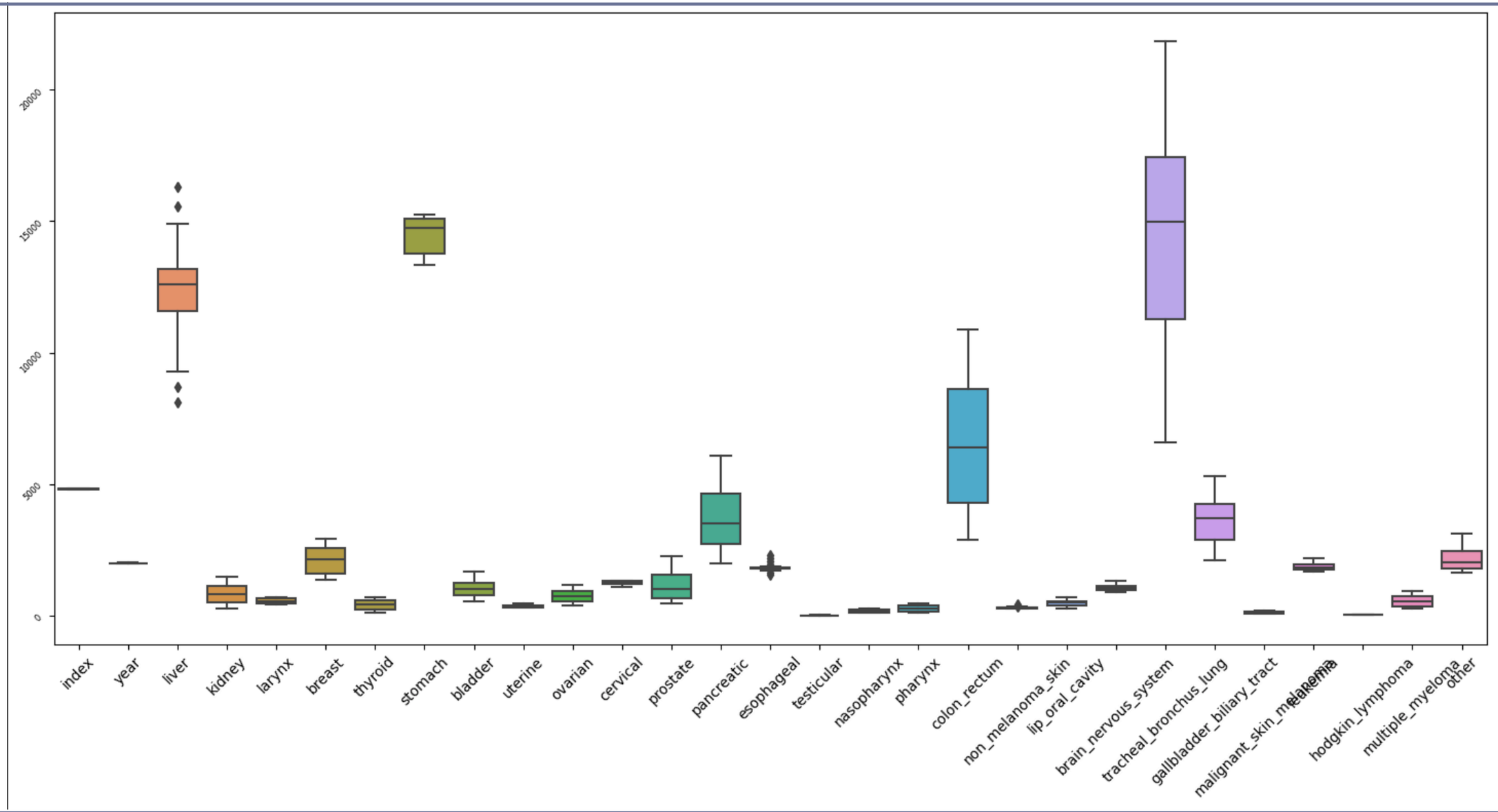

All Cancers (C00-C96)
Breast (C50)
Liver (C22)
Lung (C33-C34)
Colon and rectum (C18-C20)
Pancreas (C25)
Thyroid (C73)
Prostate (C61)
Other and unspecified (Re. C00-C96)
Stomach (C16)

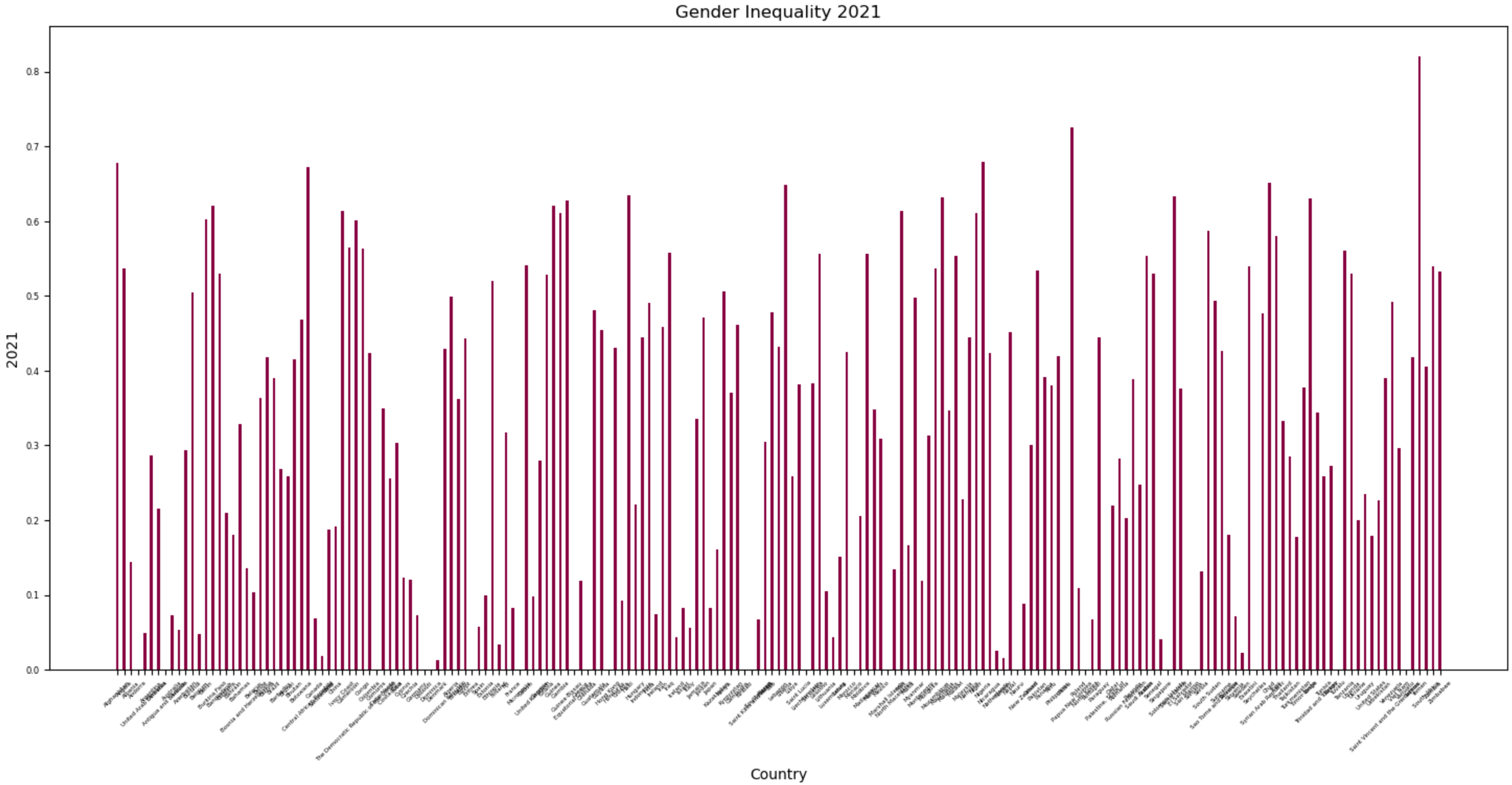
Tree02

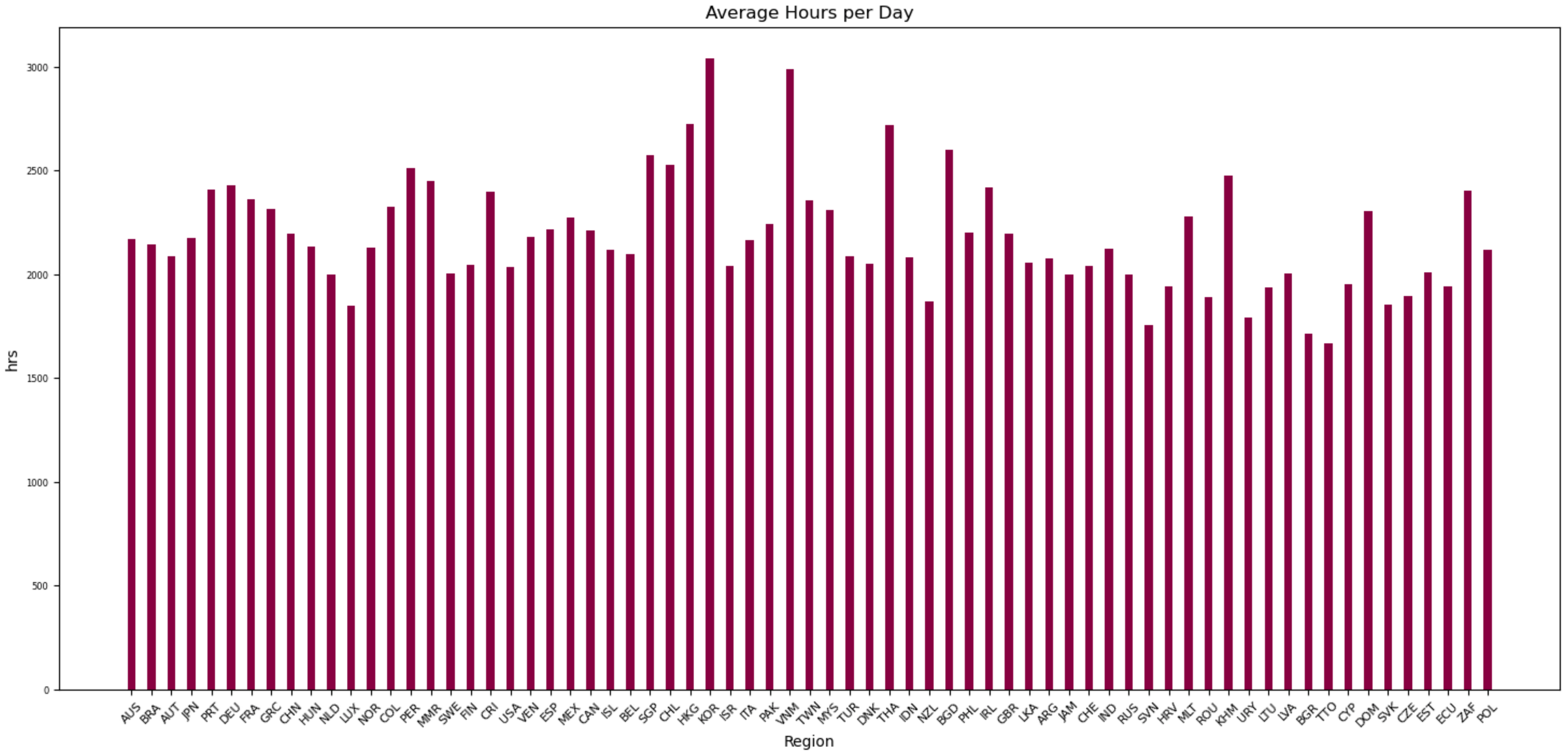

World
China
South Asia
Western Europe
Eastern Europe
Latin America and Caribbean
North Africa and
India
Tropical Latin
Brazil
Italy
East Asia
Japan
Central Latin
Ukraine
United States
Russia

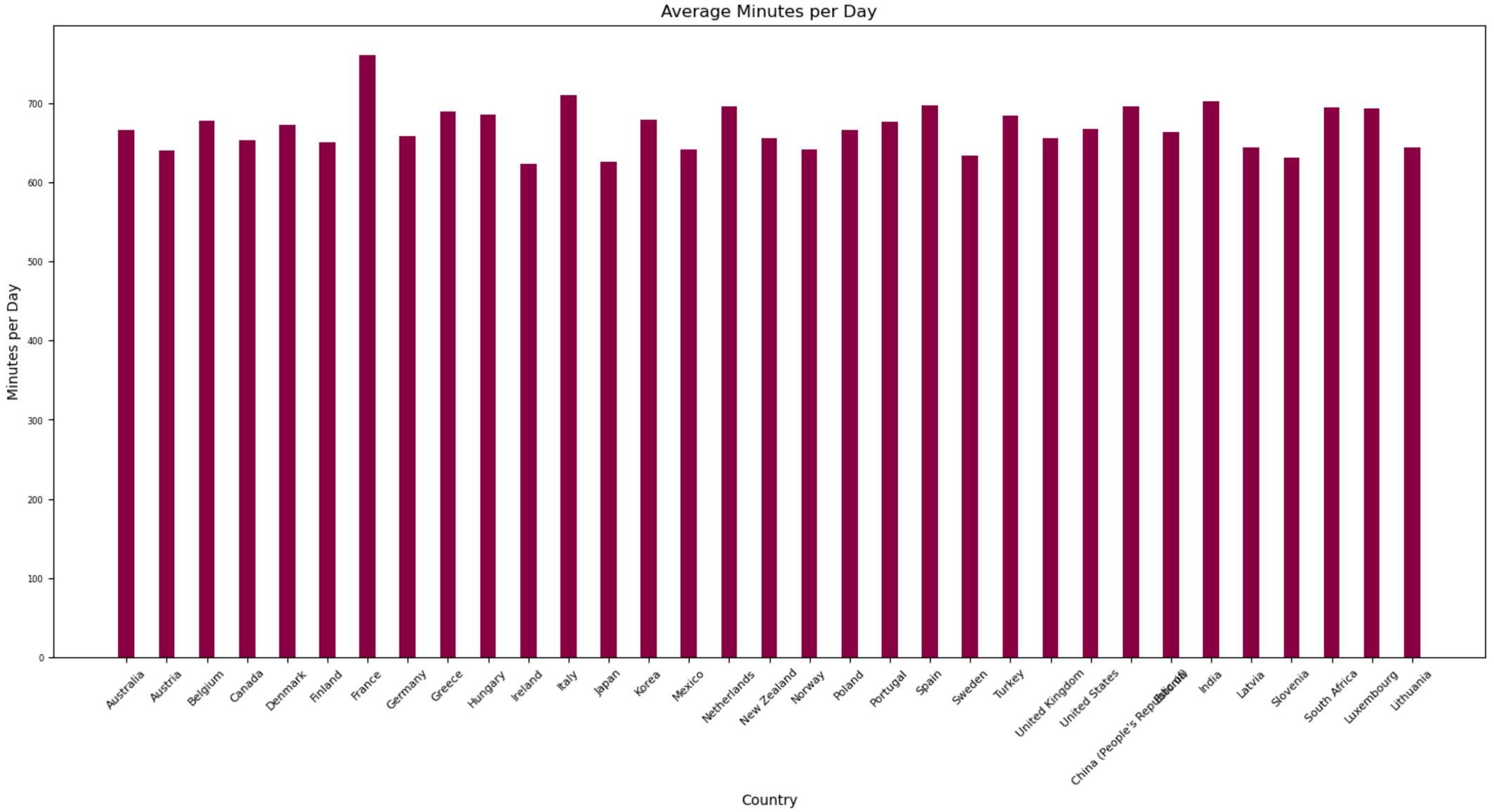
Tree02

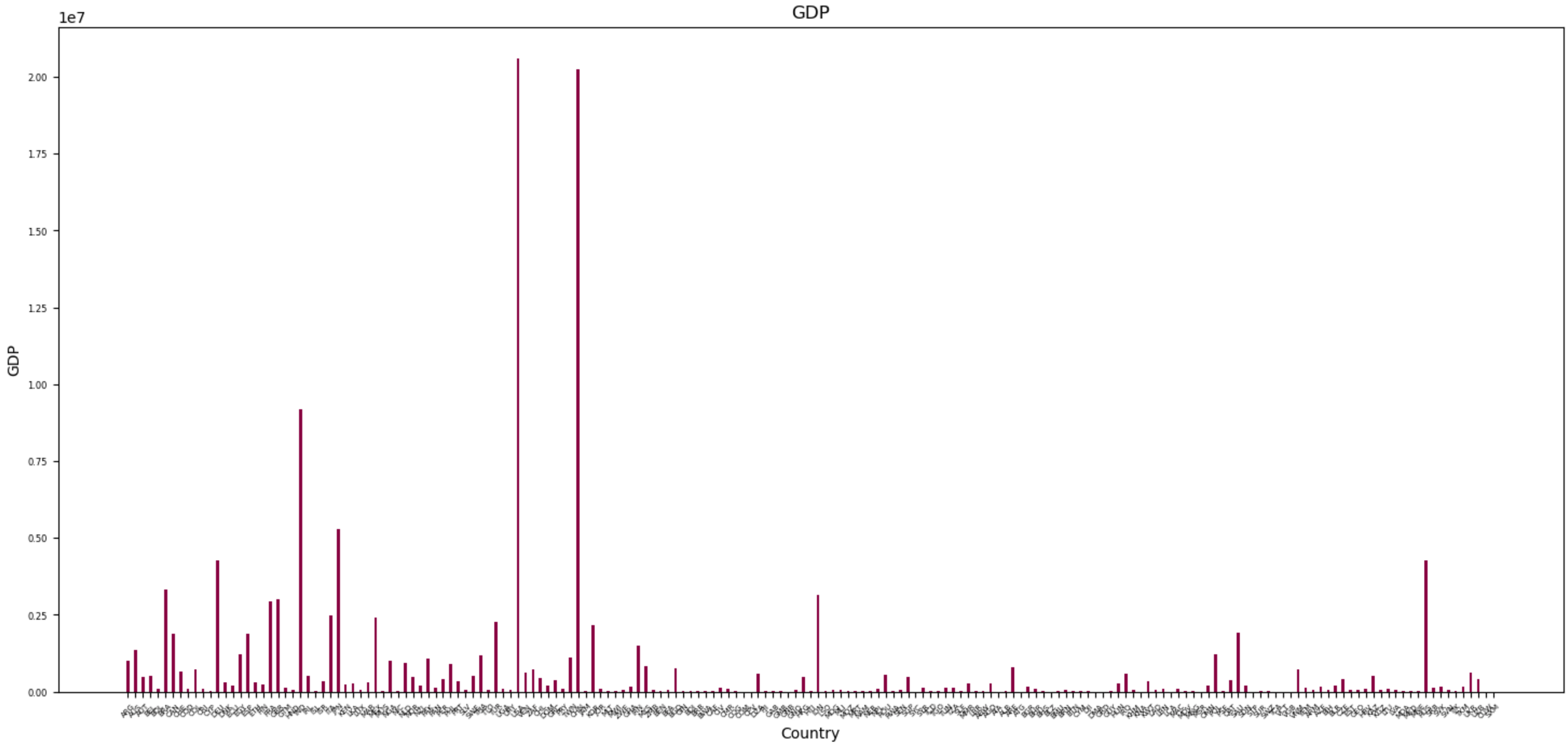

East Asia
World
Japan
China
Brazil
India
Eastern Europe
Italy
Russia
South Asia
Latin America
Western Europe


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

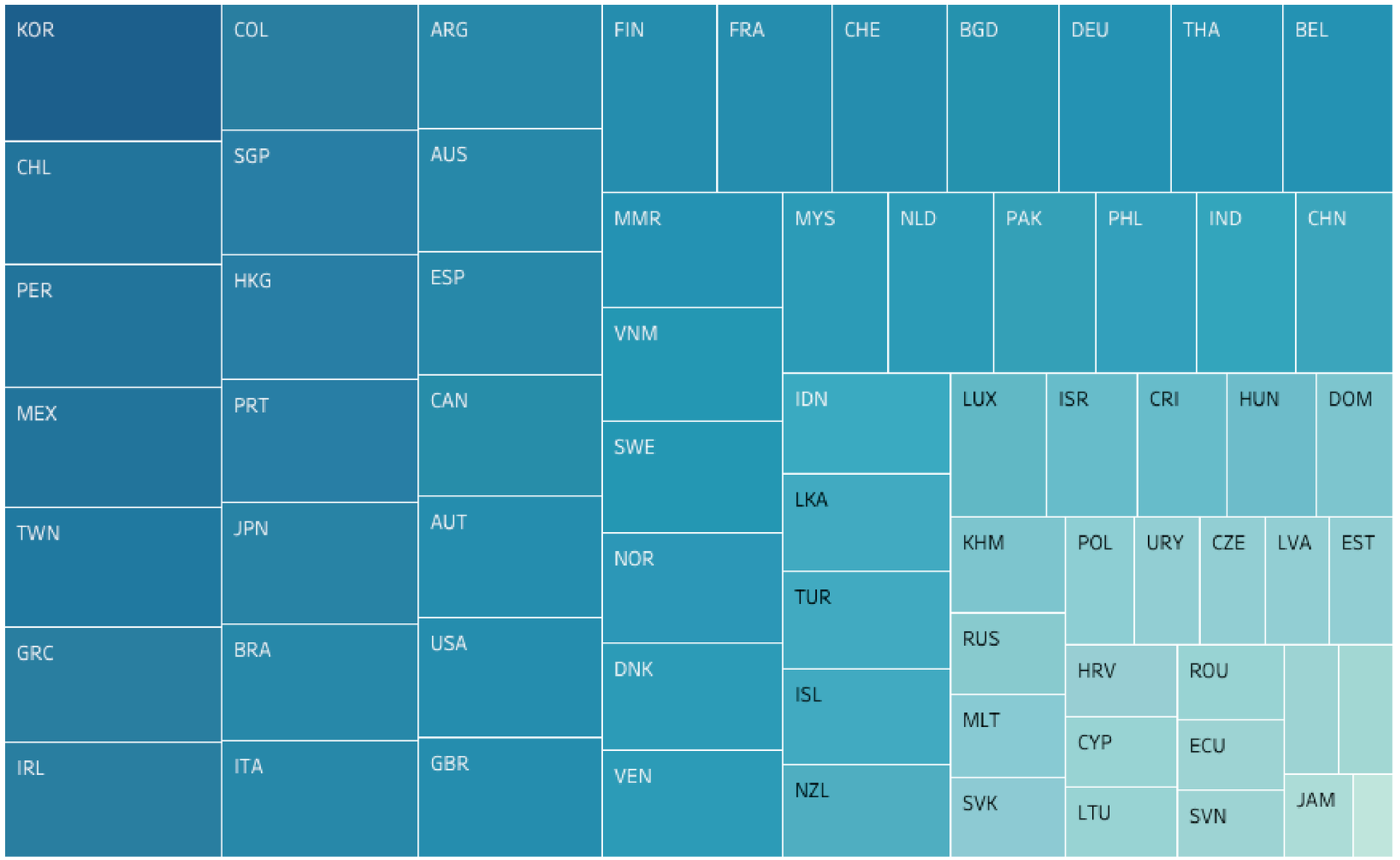

Gender Inequality 2021
2021
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Country


Average Hours per Day
hrs
0
500
1000
1500
2000
2500
3000
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 LKA ARG JAM CHE IND RUS SVN HRV MLT ROU KHM URY LTU LVA BGR TTO CYP DOM SVK CZE EST ECU ZAF POL
Region

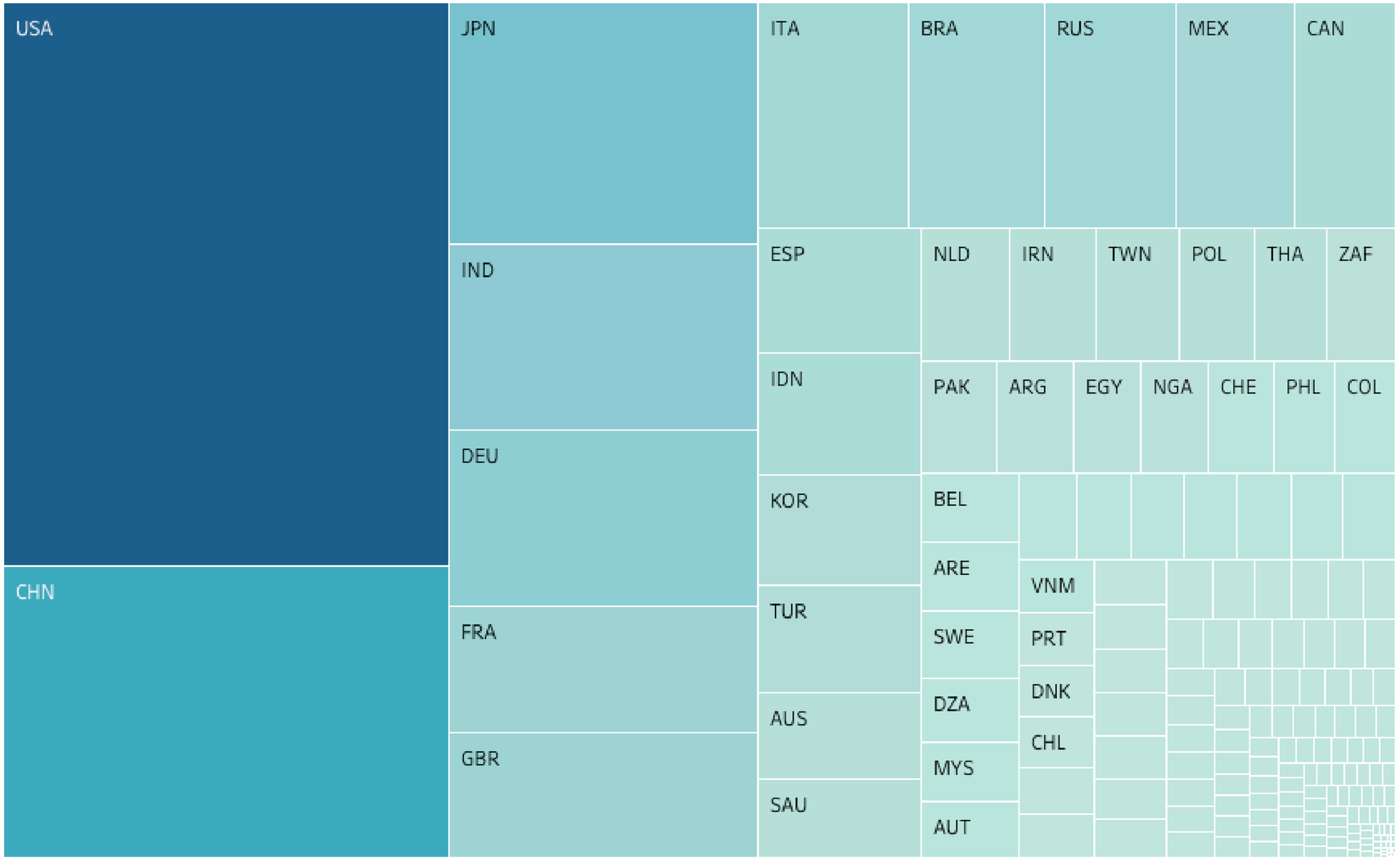

Average Minutes per Day
Minutes per Day
0
100
200
300
400
500
600
700
Australia
Austria
Belgium
Canada
Denmark
Finland
France
Germany
Greece
Hungary
Ireland
Italy
Japan
Korea
Mexico
Netherlands
New Zealand
Norway
Poland
Portugal
Spain
Sweden
Turkey
United Kingdom
United States
India
Latvia
Slovenia
South Africa
Luxembourg
Lithuania
Country


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

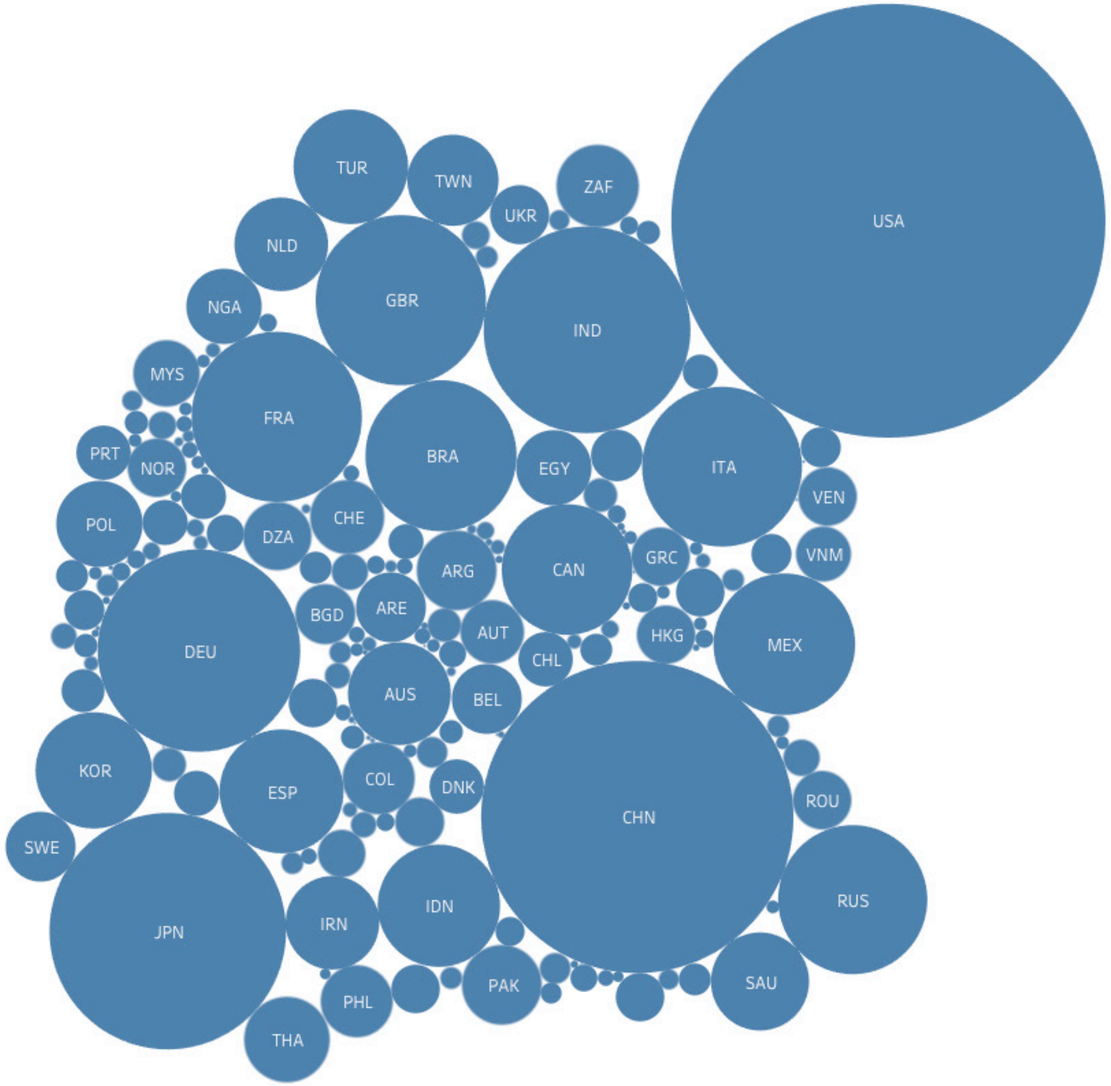
Tree


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

Tree02


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

Bubble


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

```
# 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')
```

Thank You