

DATA 607 Preparing Datasets

David Simbandumwe

2021-10-03

Introduction

The goal of this assignment is to practice preparing different datasets for downstream analysis work. For the purposes of this assignment I choose the following datasets * Global Child Mortality Rates - posted by Alec McCabe * COVID-19 Mortality Rates in NYC - William Aiken * Annual% GDP Growth - ChunJie Nan * I also added a world bank dataset that includes additional descriptive variables about a country

Global Child Mortality Rates - posted by Alec McCabe

Dataset found here: <https://sejdemyr.github.io/r-tutorials/basics/wide-and-long/>

This dataset includes child-under-5 mortality rates for all countries from 1950 to 2015. The data is structured in wide format, where the column names include the country, and each year from 1950 to 2015. Values are the corresponding child mortality rates for that country, and that year. Restructuring this dataset into long format should be very easy to accomplish with the `tidyr::gather()`.

This dataset would be a great starting point for analyzing mortality rates for children under 5 over time, by country. It would also be interesting to see if any country mortality rates are correlated over time. Monitoring spikes for mortality rate over time would be a good way to identify patterns or factors leading to child mortality.

Load Data

Loaded the data from a csv file

Tidty Data

For this analysis i wanted to evaluate the mortality data from several countries individually and grouped by region and income level.

Steps * clean the country name field - the mortality data does not include country code so we need a consistent nameing pattern to connect it with the other datasets * join the mortality data set with the combined meta * gathered the year data into rows * updated the year column to numeric * drop all rows with invalid data for u5mr

```

replace <- c("Antigua & Barbuda","Bosnia & Herzegovina","Brunei","Congo","Congo DR","Gambia The","Iran

with_this <- c("Antigua and Barbuda","Bosnia and Herzegovina","Brunei Darussalam","Congo, Rep.","Congo,

mortality_df$CountryName <- mapvalues(mortality_df$CountryName, replace,with_this)

mortality_df <- mortality_df %>%
  left_join(meta, by = c("CountryName" = "country_name")) %>%
  gather(year, u5mr, "U5MR 1950":"U5MR 2015") %>%
  mutate(year = as.numeric(gsub("U5MR.", "", year))) %>%
  drop_na(u5mr)

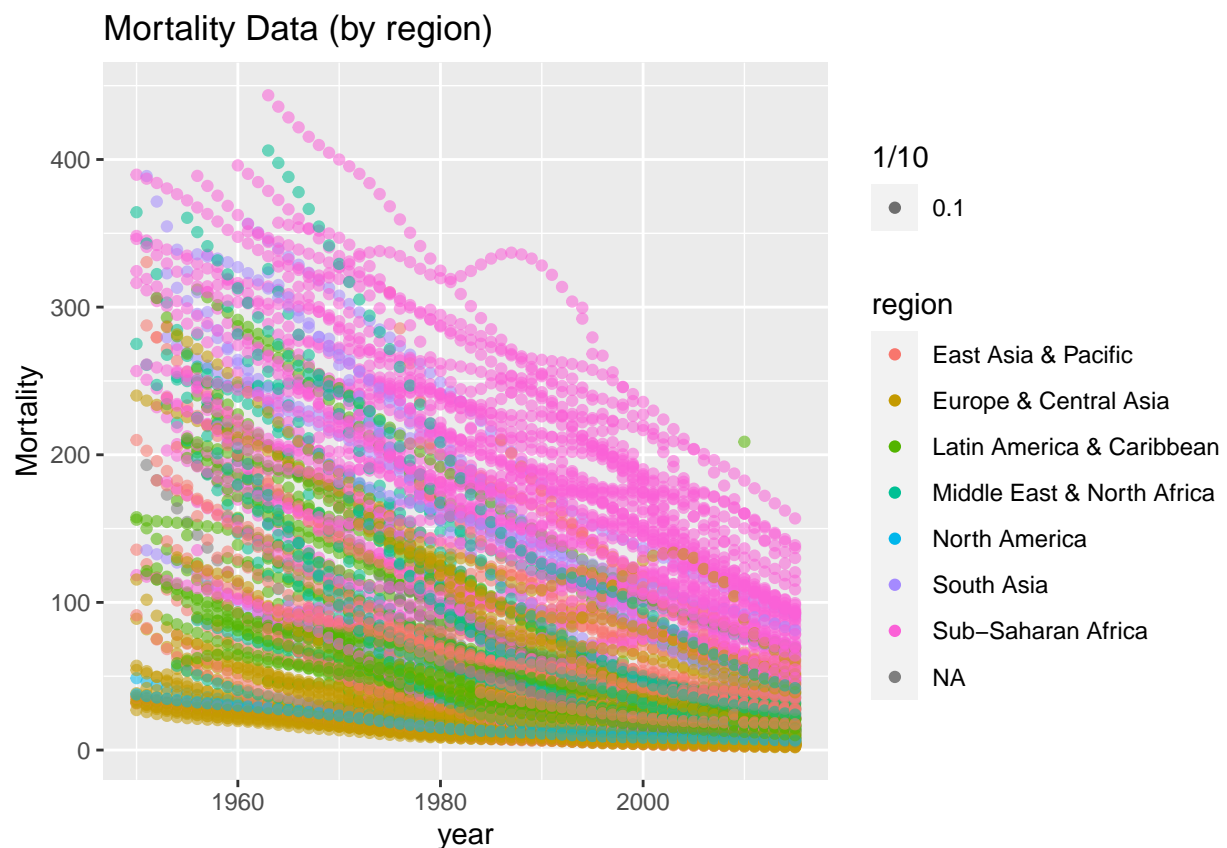
```

Analysis

```

ggplot(data = mortality_df) +
  geom_point(mapping = aes(x=year, y=u5mr, color=region, alpha=1/10)) +
  labs(
    title = "Mortality Data (by region)",
    x = "year",
    y = "Mortality"
  )

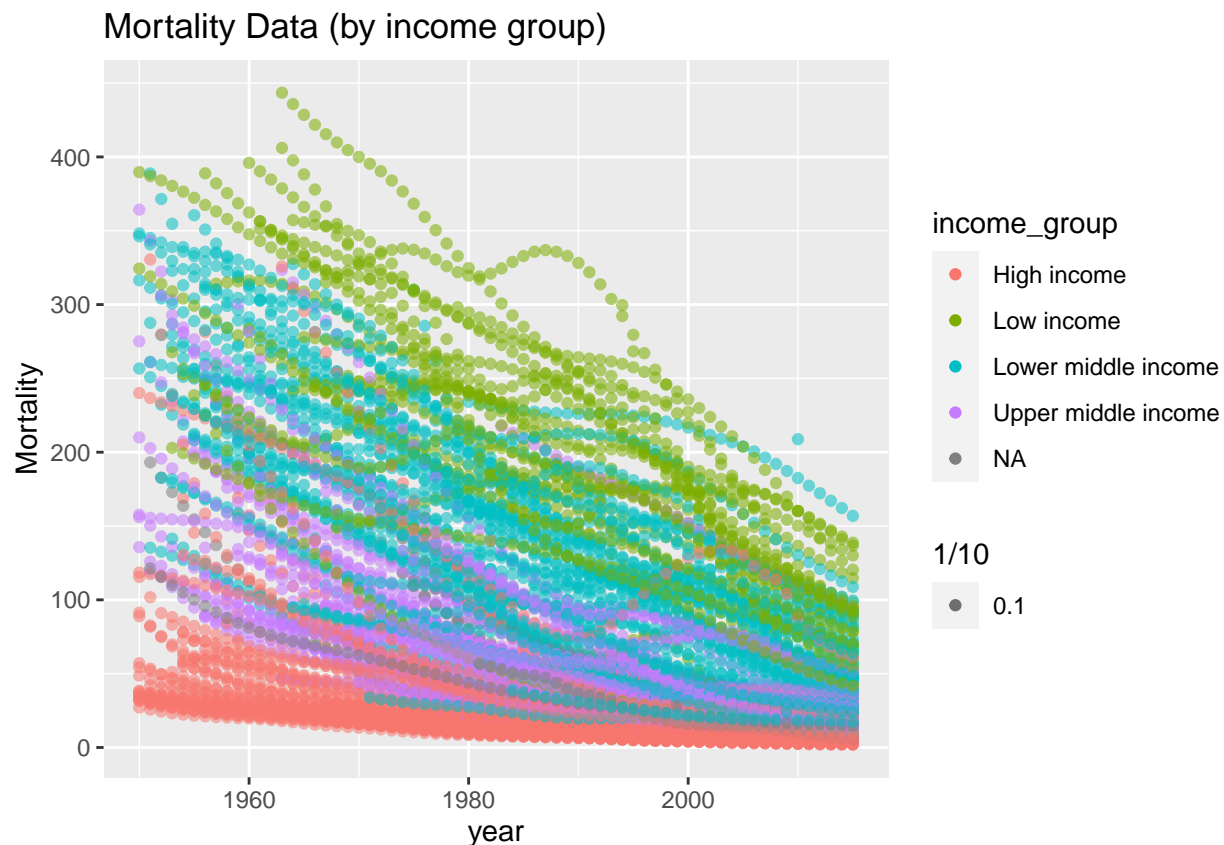
```



There are large drops in the mortality rates across all countries from 1950 to 2015. However there still

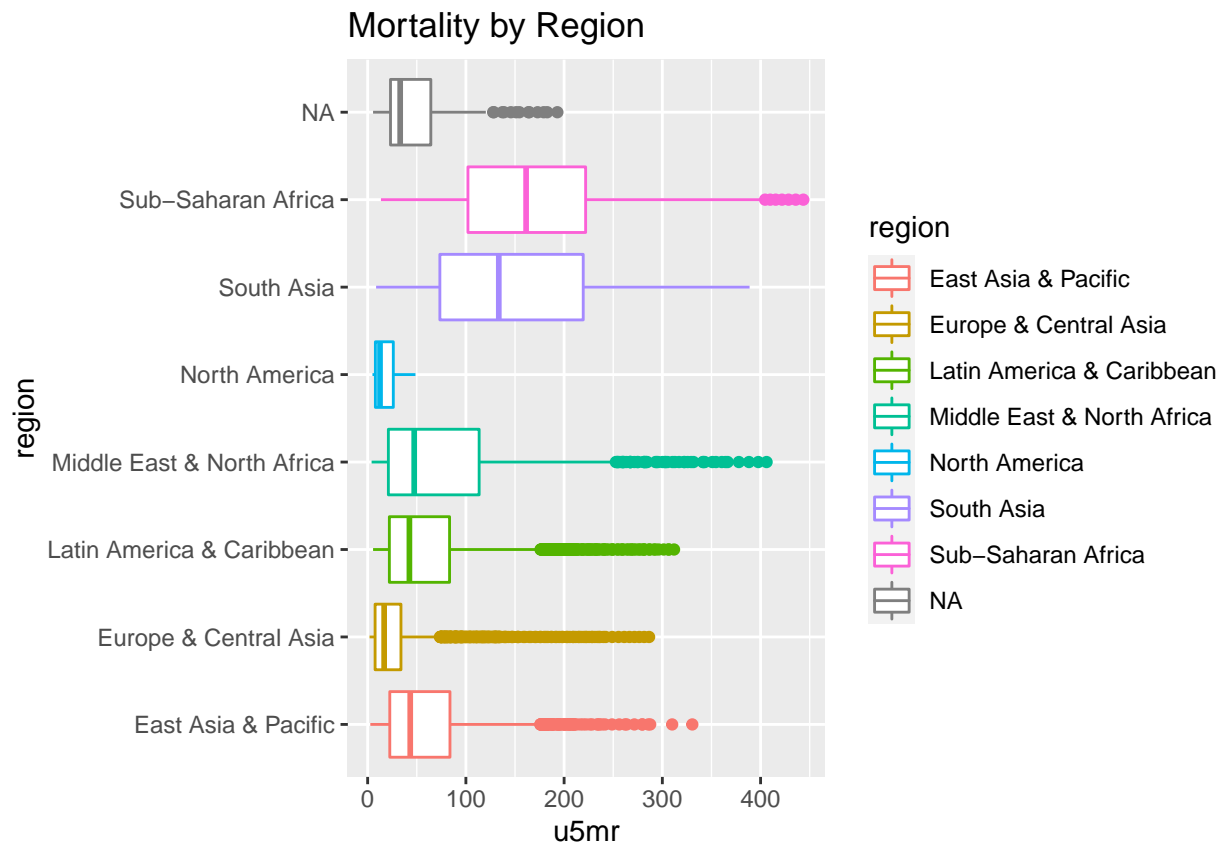
remains a substantial differential in the rates from European and North American countries. The difference between the regions has been shrinking over the years.

```
ggplot(data = mortality_df) +
  geom_point(mapping = aes(x=year, y=u5mr, color=income_group, alpha=1/10)) +
  labs(
    title = "Mortality Data (by income group)",
    x = "year",
    y = "Mortality"
  )
```



We see similar results when we color the data by income groups. General declines across all countries with a pronounced difference between the high and the low income countries.

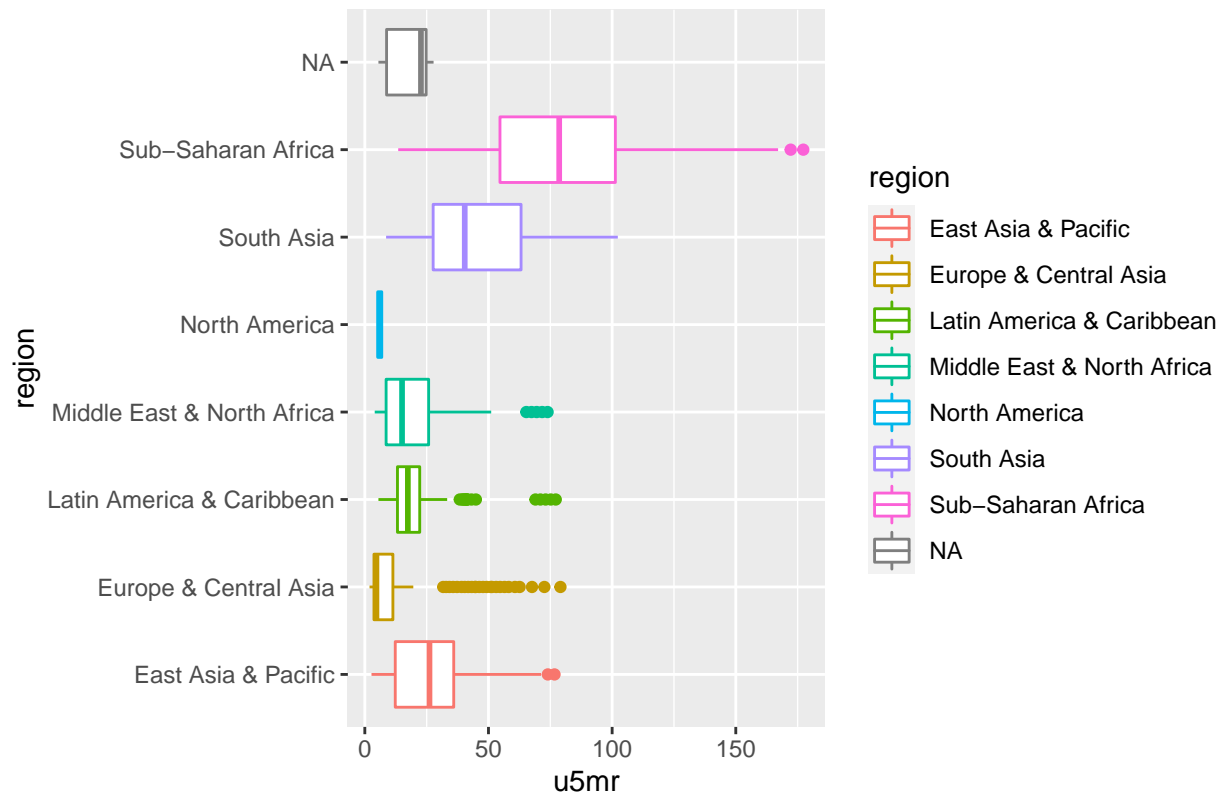
```
ggplot(data = mortality_df) +
  geom_boxplot(mapping = aes(y=region, x=u5mr, color=region)) +
  labs (
    title = "Mortality by Region"
  )
```



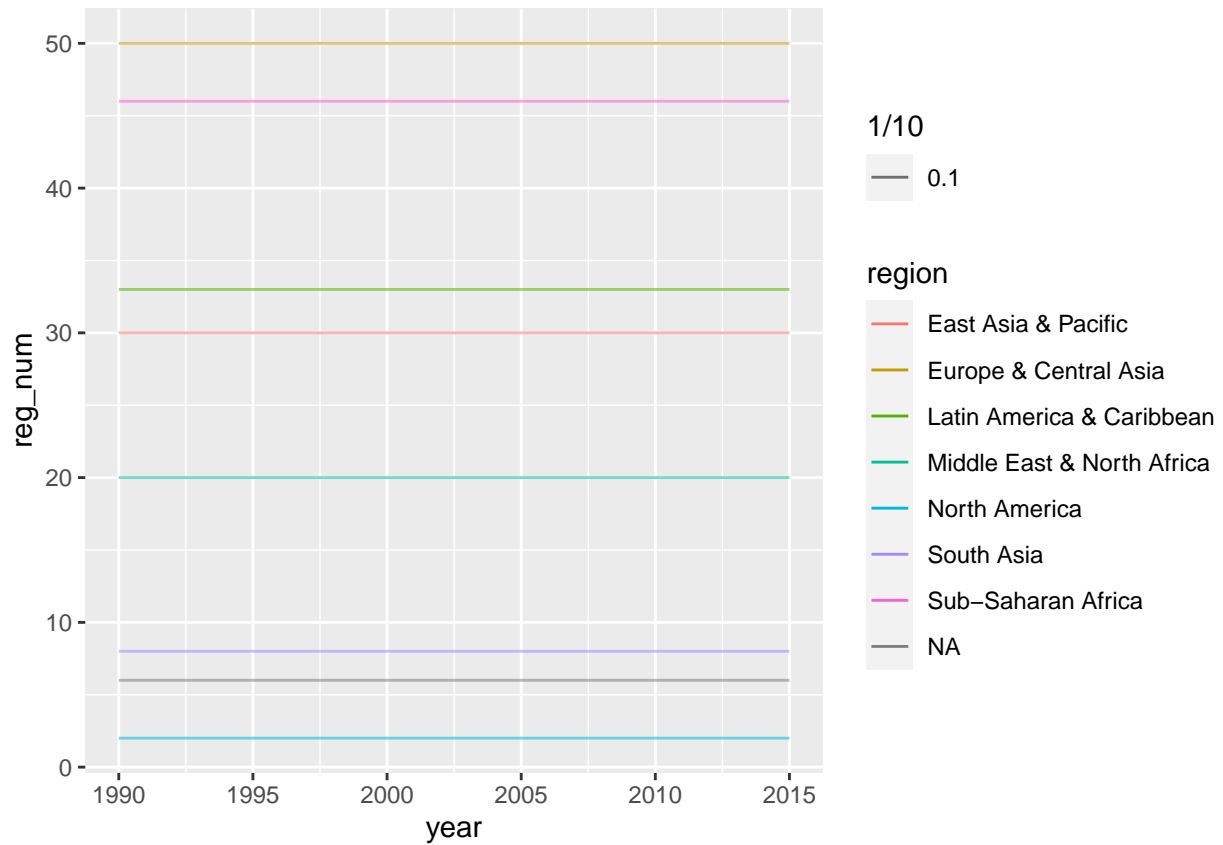
To view the data slightly differently highlighting the variation by region we can view a box plot of the full data set. Sub-Saharan Africa, South Asia, the Middle East & North Africa show larger interquartile range.

```
mortality_df %>%
  filter(year > 2010) %>%
  ggplot() +
    geom_boxplot(mapping = aes(y=region, x=u5mr, color=region)) +
    labs (
      title = "Mortality by Region (2010 - 2015)"
    )
```

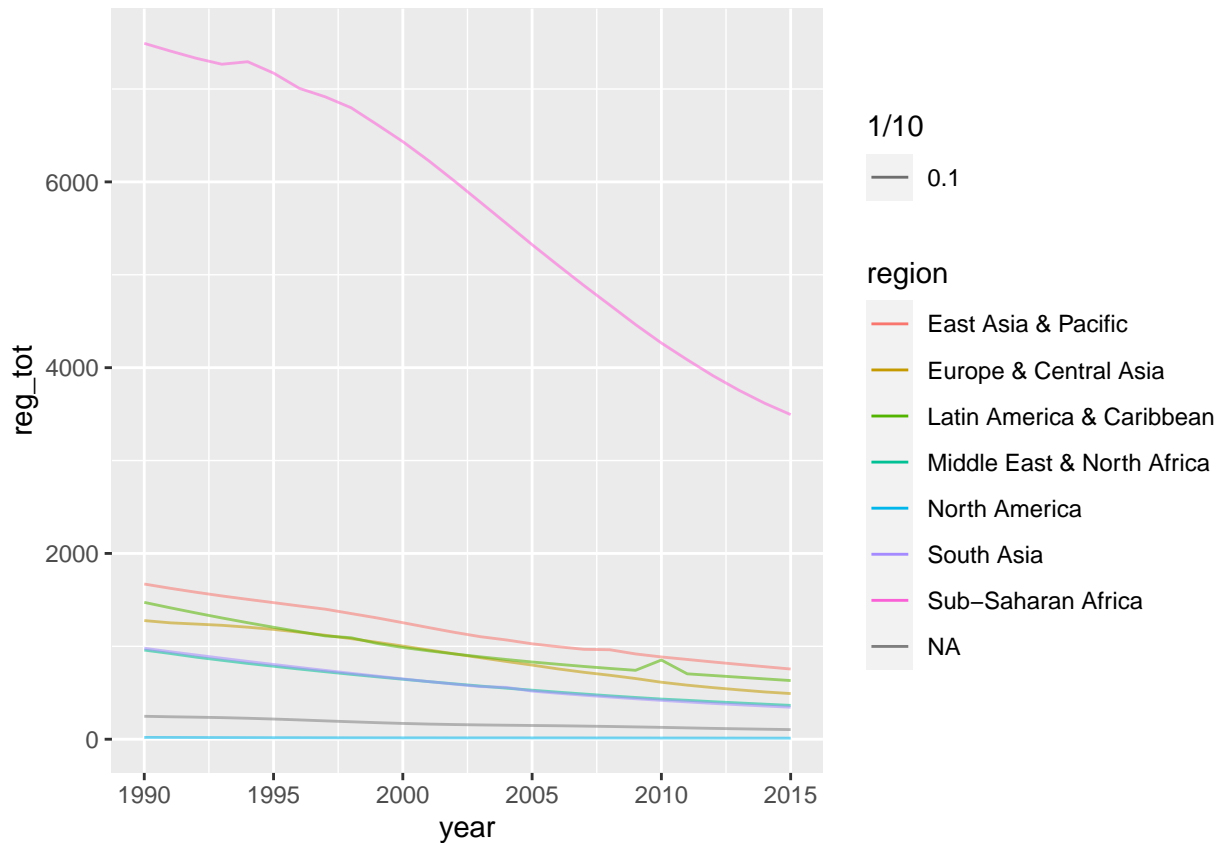
Mortality by Region (2010 – 2015)



```
mortality_df %>%
  filter(year > 1989) %>%
  group_by(year, region) %>%
  dplyr::mutate(
    reg_tot = sum(u5mr),
    reg_num = n()
  ) %>%
  select(region, year, reg_tot, reg_num) %>%
  distinct() %>%
  ggplot() +
  geom_line(mapping = aes(x=year, y=reg_num, color=region, alpha=1/10))
```



```
mortality_df %>%
  filter(year > 1989) %>%
  group_by(year, region) %>%
  dplyr::mutate(
    reg_tot = sum(u5mr),
    reg_num = n()
  ) %>%
  select(region, year, reg_tot, reg_num) %>%
  distinct() %>%
  ggplot() +
  geom_line(mapping = aes(x=year, y=reg_tot, color=region, alpha=1/10))
```



The regional view of the data shows a similar pattern. We adjusted the timeframe to include only the years where we have full records per region. As whole each region shows decreased child mortality overtime however Sub-Saharan Africa is trailing

Conclusion

All of the analysis points to the same conclusions. Child mortality has improved over the years however there still remains clear distinctions in mortality levels based on region and income.

COVID-19 Mortality Rates in NYC

The nyc.gov website has both the hospitalization and mortality rates by county and zip code over time. It would be interesting to compare the relationship between these two data set. How it hospitalization and death rates differ by county in NYC?

<https://www1.nyc.gov/site/doh/covid/covid-19-data-totals.page>

Load Data

Load data from a csv file

Tiddy Data

For this analysis I wanted to focus on deaths and hospitalization. Removed the unneeded columns and Transformed the wide data into tall data set.

Steps: * removed the columns that are not needed * gather the columns data to transform the wide dataset into a long dataset * create 2 variable out of the resulting key. one for borough and the other for the type of event * drop the n/a in the count field * clean up the column names

```
covid_df <- covid_df %>%
  select (
    group,
    subgroup,
    BK_HOSPITALIZED_COUNT,
    BK_DEATH_COUNT,
    BX_HOSPITALIZED_COUNT,
    BX_DEATH_COUNT,
    MN_HOSPITALIZED_COUNT,
    MN_DEATH_COUNT,
    QN_HOSPITALIZED_COUNT,
    QN_DEATH_COUNT,
    SI_HOSPITALIZED_COUNT,
    SI_DEATH_COUNT
  ) %>%
  gather(type, count,
    "BK_HOSPITALIZED_COUNT",
    "BK_DEATH_COUNT",
    "BX_HOSPITALIZED_COUNT",
    "BX_DEATH_COUNT",
    "MN_HOSPITALIZED_COUNT",
    "MN_DEATH_COUNT",
    "QN_HOSPITALIZED_COUNT",
    "QN_DEATH_COUNT",
    "SI_HOSPITALIZED_COUNT",
    "SI_DEATH_COUNT"
  ) %>%
  separate(
    type,
    into = c("borough", "event", "junk"),
    extra = "merge",
    fill = "left",
    convert = TRUE,
    sep = "\\_"
  ) %>%
  drop_na(count)

# drop junk column
covid_df <- covid_df[ , !(names(covid_df) == "junk")]
```

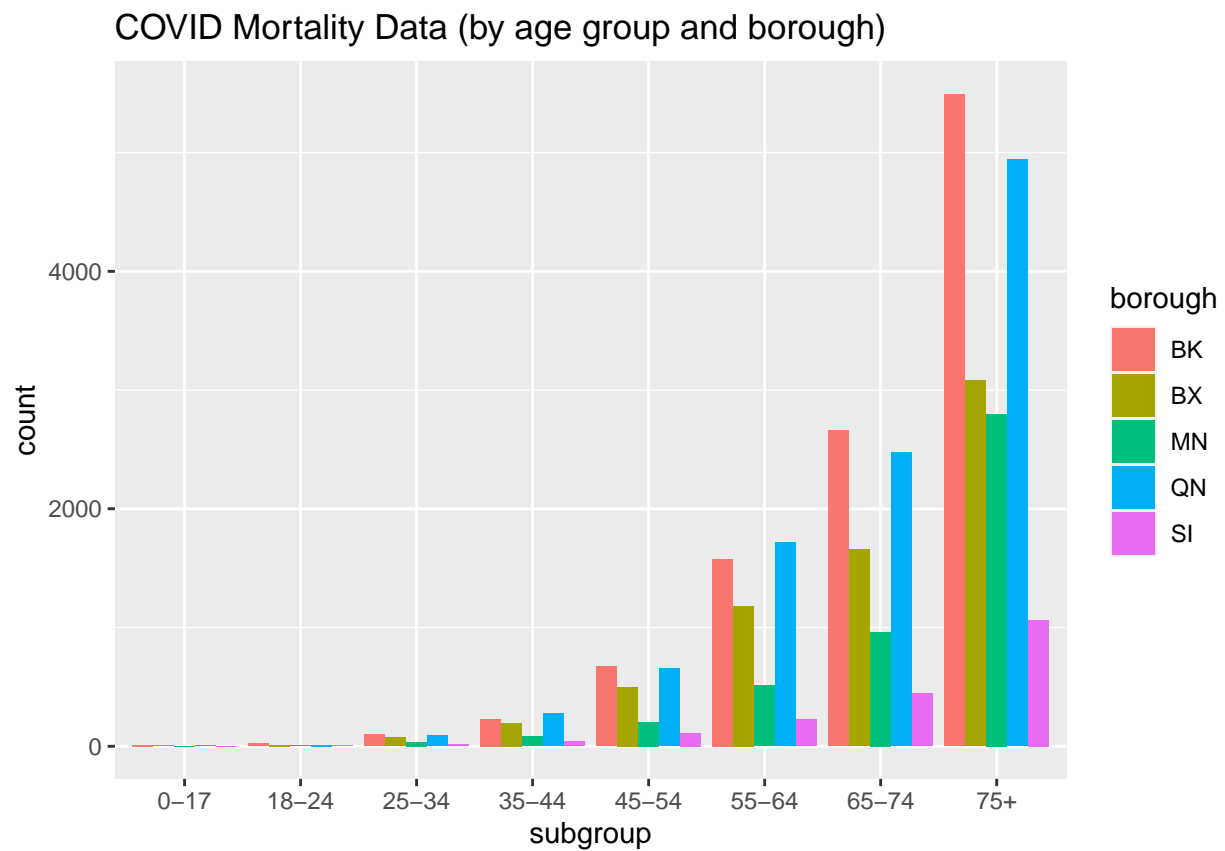
```
covid_df %>%
  filter(group == "Age", event=="DEATH") %>%
  ggplot(aes(fill=borough, y=count, x=subgroup)) +
  geom_bar(position="dodge", stat="identity") +
  labs(
```



```

    title = "COVID Mortality Data (by age group and borough)"
  )

```

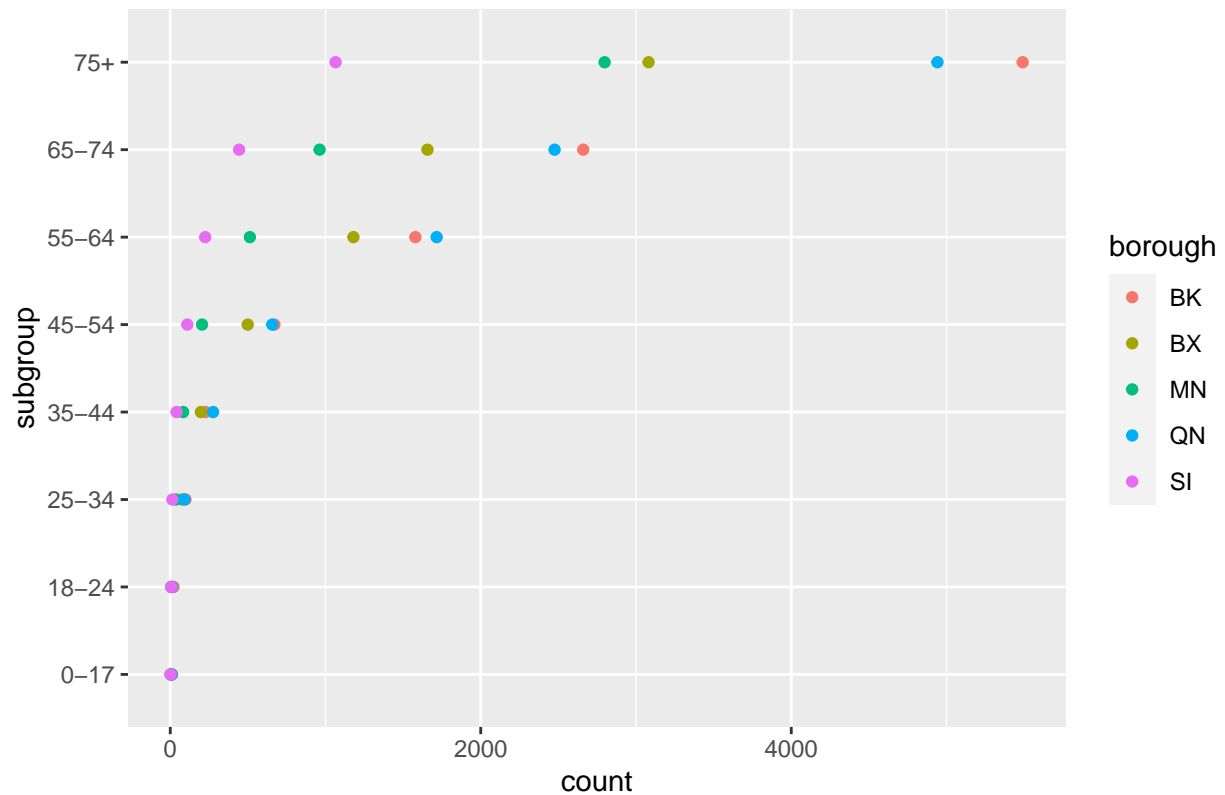


```

covid_df %>%
  filter(group == "Age", event=="DEATH") %>%
  ggplot() +
    geom_point(mapping = aes(x=count, y=subgroup, color=borough)) +
    labs(
      title = "COVID Mortality Data (by age group and borough)"
    )

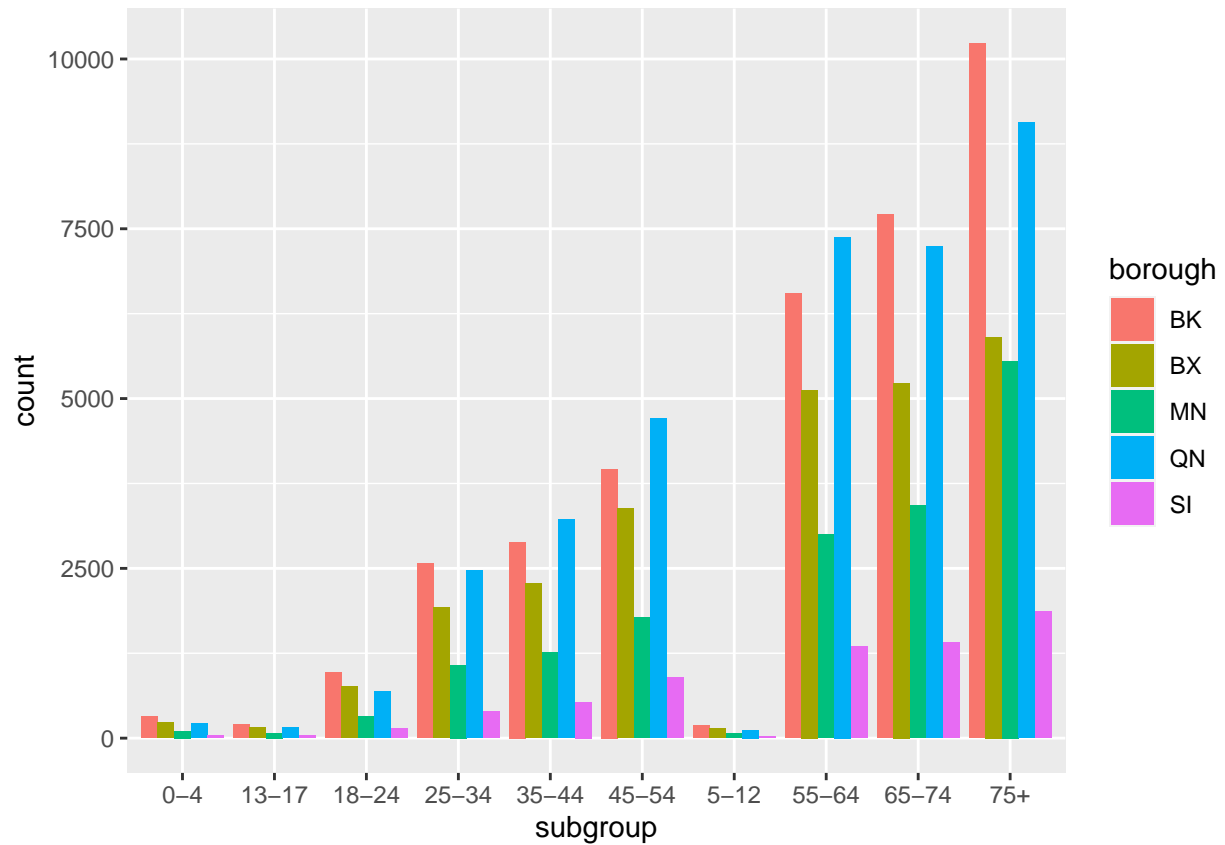
```

COVID Mortality Data (by age group and borough)



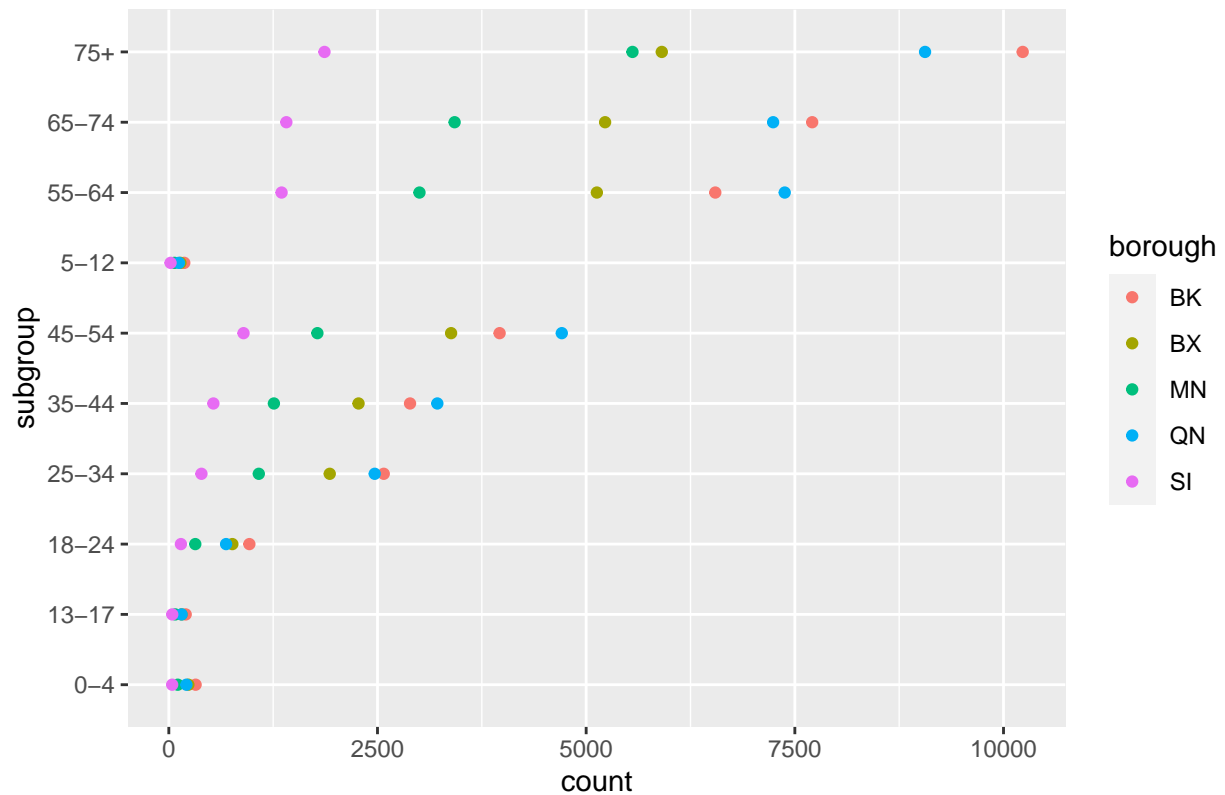
The age, mortality plot highlight the impact of borough population and age on the number of deaths from covid. We can see boroughs with larger populations see increases in total deaths for older residence.

```
covid_df %>%
  filter(group == "Age", event=="HOSPITALIZED") %>%
  ggplot(aes(fill=borough, y=count, x=subgroup)) +
  geom_bar(position="dodge", stat="identity")
```



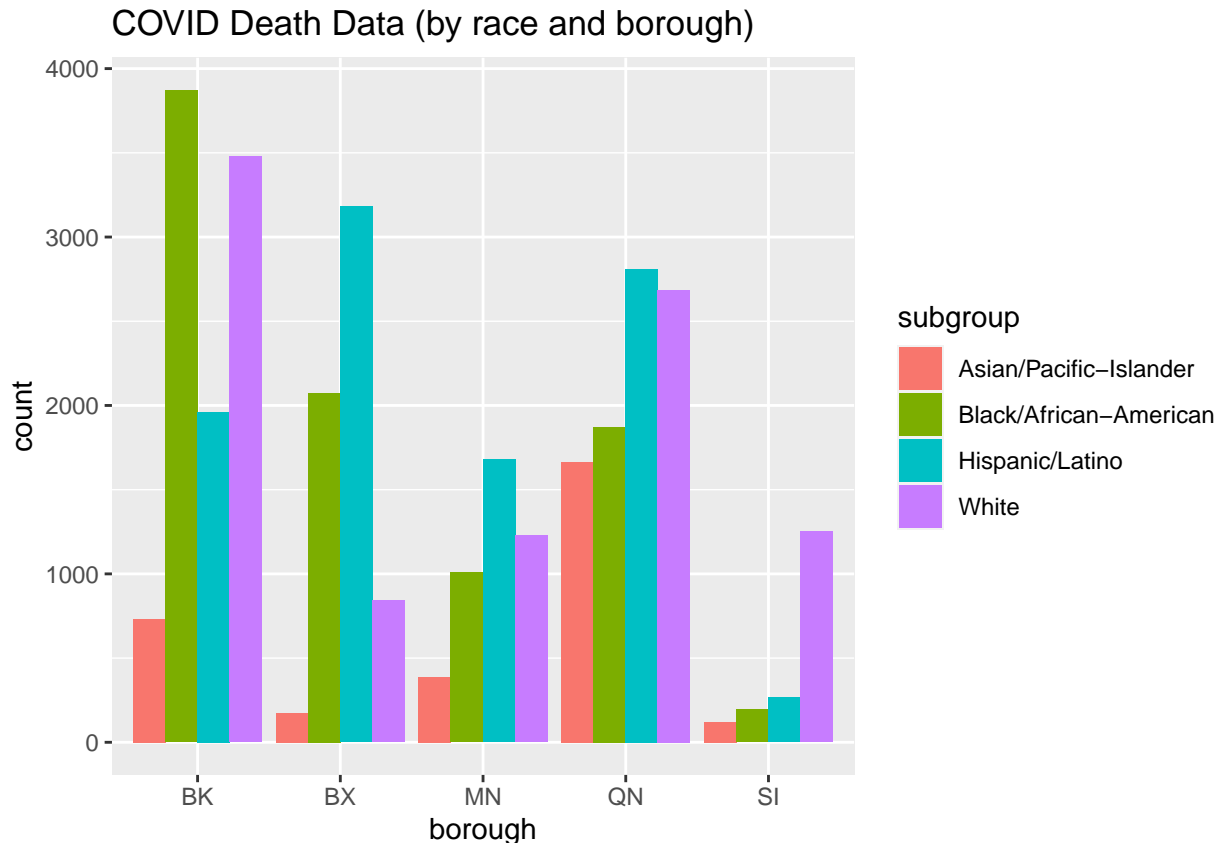
```
covid_df %>%
  filter(group == "Age", event=="HOSPITALIZED") %>%
  ggplot() +
  geom_point(mapping = aes(x=count, y=subgroup, color=borough)) +
  labs(
    title = "COVID Hospitalizations Data (by income age group and borough)"
  )
```

COVID Hospitalizations Data (by income age group and borough)



The same trend observed with deaths is slightly let pronounced with hospitalization but it still exists.

```
covid_df %>%
  filter(group == "Race/ethnicity", event=="DEATH") %>%
  ggplot(aes(fill=subgroup, y=count, x=borough)) +
  geom_bar(position="dodge", stat="identity") +
  labs(
    title = "COVID Death Data (by race and borough)"
  )
```



The grouped bar chart shows some difference in total deaths by race for each borough. For follow-up analysis, it would be good to look at these numbers as a percentage of the population.

Conclusion

This analysis would be more impactful if we reviewed the data using the relative sizes of each population to determine if specific age groups or demographic groups are overrepresented in the counts for hospitalizations and deaths.

Annual% GDP Growth - ChunJie Nan

The GDP growth(annual%) is a data from The World Bank. It includes 266 observations/countries' % annual GDP growth. The data set is tremendous and has some NA/missing values.

It requires some data adjustment, such as handling the missing value with zoo package and subset the data to a small group especially the top 5 GDP countries to see which high GDP countries has

affected the most by the Covid-19. With the historical data, we can forecast/predict the year 2020 GDP from the difference between the real value GDP of 2020 and the predicted value of the year

2020 GDP, I can find out how much does Covid-19 affected the GDP growth.

data source:

<https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG>

Load Data

Load data from a csv file. There is some additional file information at the start of the file so the first line in the file is not a valid list of column names.

Tidty Data

I wanted to analyze the gdp data for each country, regional and income group so i have joined the original dataset with a world bank dataset that incldues additional attributes for each country. Also is a similar dataset from the first part of the assignment so we can combine the two data sets to see if there is any correlation.

The steps taken: clean and filter the world bank data * WB DATA select the columns of interest from the world bank data * WB DATA gather year columns * WB DATA clean the column names * WB DATA drop rows with na data * WB DATA remove unneeded columns * WB DATA transform column data into numerics * WB DATA spread the world bank data by value * WB_DATA rename columns

The steps taken: clean and filter the gdp data * gather the wide year data into rows * drop na data * join the gdp data with the world bank data set * clean up the rows * remove unused columns

```
c_measures <- c("SP.POP.TOTL", "SP.POP.GROW", "SP.DYN.LE00.IN", "SP.DYN.TFRT.IN", "SP.ADO.TFRT", "SH.STA.MALN.ZS", "SH.IMM.MEAS", "SE.SEC.ENRR", "SE.ENR.PRSC.FM.ZS", "EG.USE.PCAP.KG.OE", "EN.ATM.CO2E.PC")
```

```
wb_df <- wb_df %>%
  subset(series_code %in% c_measures) %>%
  gather(year, value, "2000[YR2000]":"2015[YR2015]") %>%
  mutate(year = as.numeric(gsub("\\[YR20\\d+\\]", "", year))) %>%
  drop_na(value) %>%
  subset(select = -series_name )
```

```
wb_df$value = as.numeric(wb_df$value)
```

```
## Warning: NAs introduced by coercion
```

```
wb_df <- wb_df %>%
  spread(series_code, value)
```

```
wb_df <- wb_df %>%
  rename(
    c("SP.POP.TOTL" = "population_total",
      "SP.POP.GROW" = "population_growth",
      "SP.DYN.LE00.IN" = "life_expectancy",
      "SP.DYN.TFRT.IN" = "fertility_rate",
      "SP.ADO.TFRT" = "adolescent_fertility_rate",
      "SH.STA.MALN.ZS" = "underweight_u5",
      "SH.IMM.MEAS" = "immunization_measles",
      "SE.SEC.ENRR" = "enrollment_secondary_school",
      "SE.ENR.PRSC.FM.ZS" = "enrollment_primary_secondary",
      "EG.USE.PCAP.KG.OE" = "energy_use",
      "EN.ATM.CO2E.PC" = "co2_emissions",
```

```

"EG.USE.ELEC.KH.PC" = "electric_power_cons",
"NY.GDP.MKTP.CD" = "gdp",
"NY.GDP.MKTP.KD.ZG" = "gdp_growth",
"NY.GDP.DEFL.KD.ZG" = "gdp_deflator",
"NV.AGR.TOTL.ZS" = "ag_value_add",
"NV.IND.TOTL.ZS" = "industry_value_add",
"NE.GDI.TOTL.ZS" = "gross_capital_formation",
"GC.REV.XGRT.GD.ZS" = "revenue_no_grants",
"CM.MKT.LCAP.GD.ZS" = "mkt_cap",
"MS.MIL.XPND.GD.ZS" = "mil_expenditures",
"IT.CEL.SETS.P2" = "cell_subs",
"TG.VAL.TOTL.GD.ZS" = "merch_trade",
"TT.PRI.MRCH.XD.WD" = "net_barter_terms",
"BM.TRF.PWKR.CD.DT" = "personal_remittance",
"BX.KLT.DINV.CD.WD" = "foreign_direct_investment",
"NY.GDP.PCAP.CD" = "gdp_per_capita",
"BN.KLT.DINV.CD" = "foreign_direct_investment_net",
"FP.CPI.TOTL.ZG" = "consumer_price_index"
)
)

gdp_df <- gdp_df %>%
  gather(year, growth, "1960":"2020") %>%
  drop_na(growth) %>%
  select(CountryName, CountryCode, year, growth) %>%
  transform(year = as.numeric(year))

comb_df <- gdp_df %>%
  inner_join(wb_df, by = c("CountryCode" = "country_code", "year")) %>%
  left_join(mortality_df, by = c("CountryCode" = "country_id", "year" )) %>%
  subset(select = -c(CountryName.x, CountryName.y) )

```

Analysis

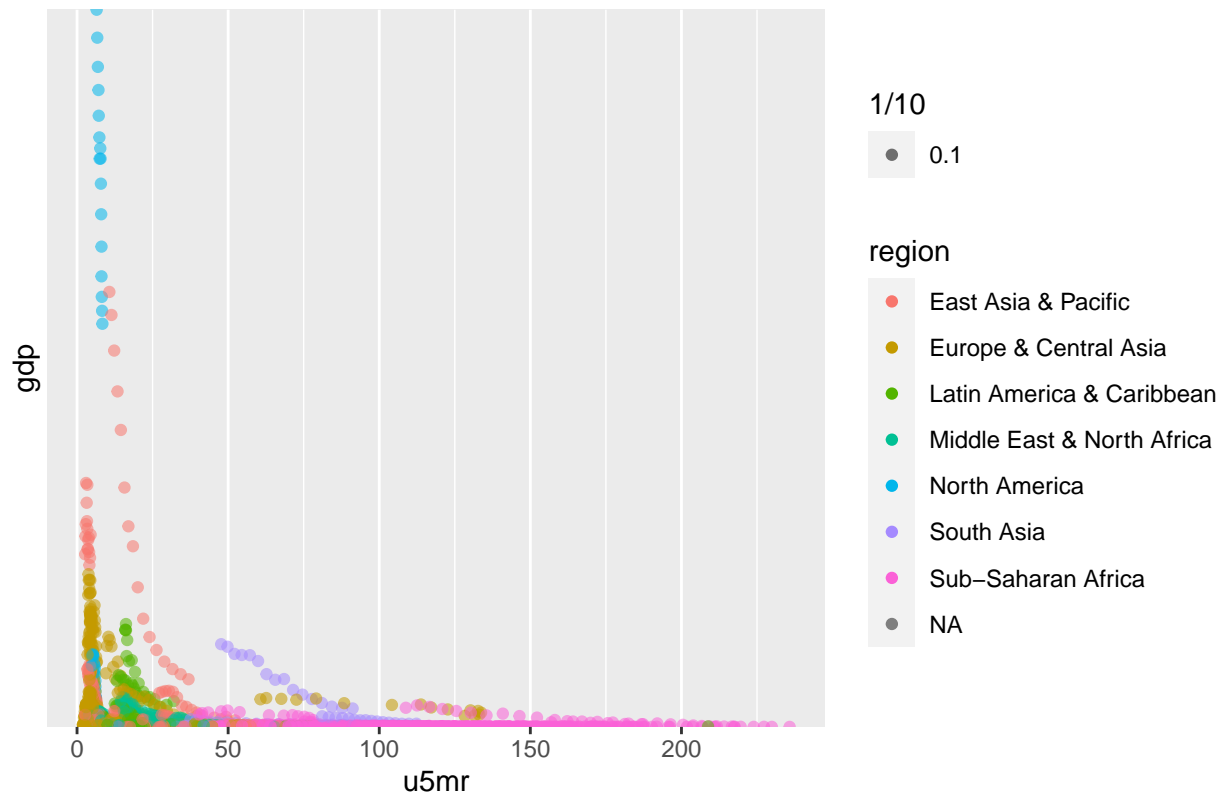
```

comb_df %>%
  ggplot() +
    geom_point(mapping = aes(y=gdp, x=u5mr, color=region, alpha=1/10)) +
    scale_y_discrete(breaks = 20) +
    labs(
      title = "GDP Compaired ot Child Mortality"
    )

```

```
## Warning: Removed 334 rows containing missing values (geom_point).
```

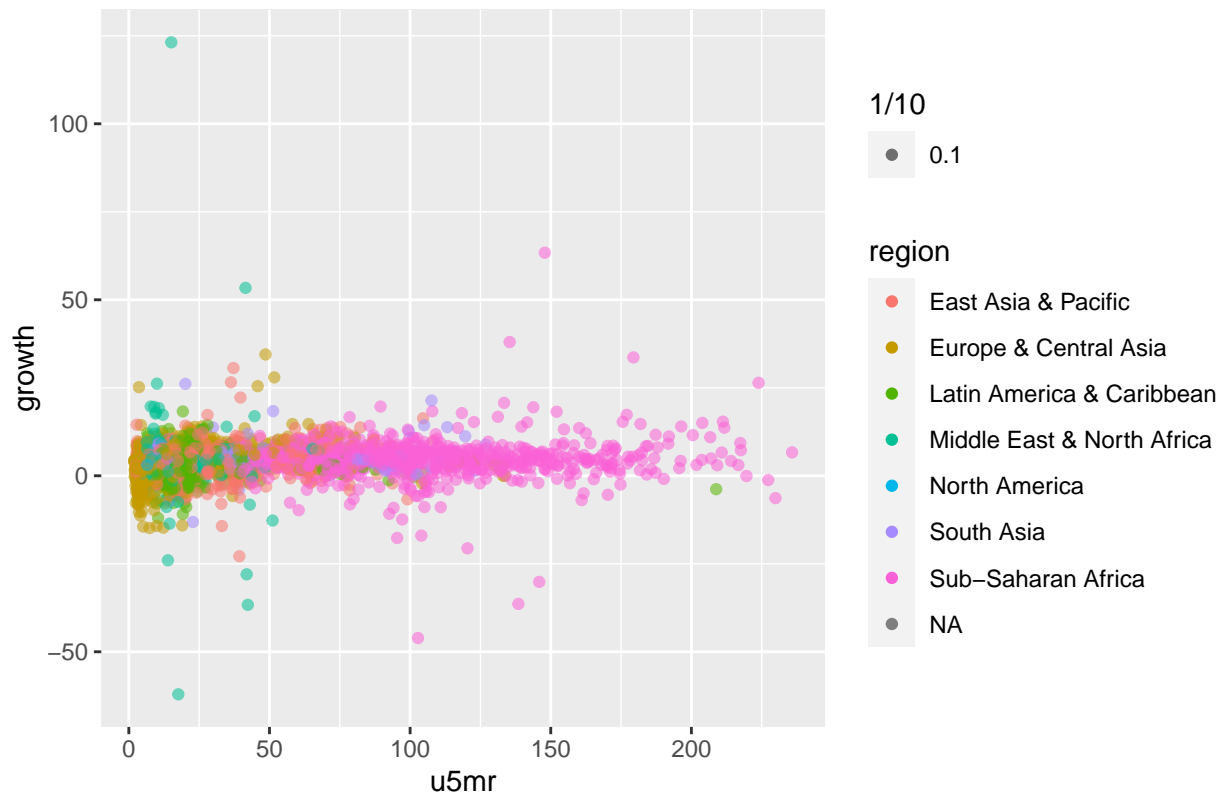
GDP Compared to Child Mortality



```
ggplot(data = comb_df) +
  geom_point(mapping = aes(x=u5mr, y=growth, color=region, alpha=1/10)) +
  labs(
    title = "GDP Growth Compared to Child Mortality"
  )
```

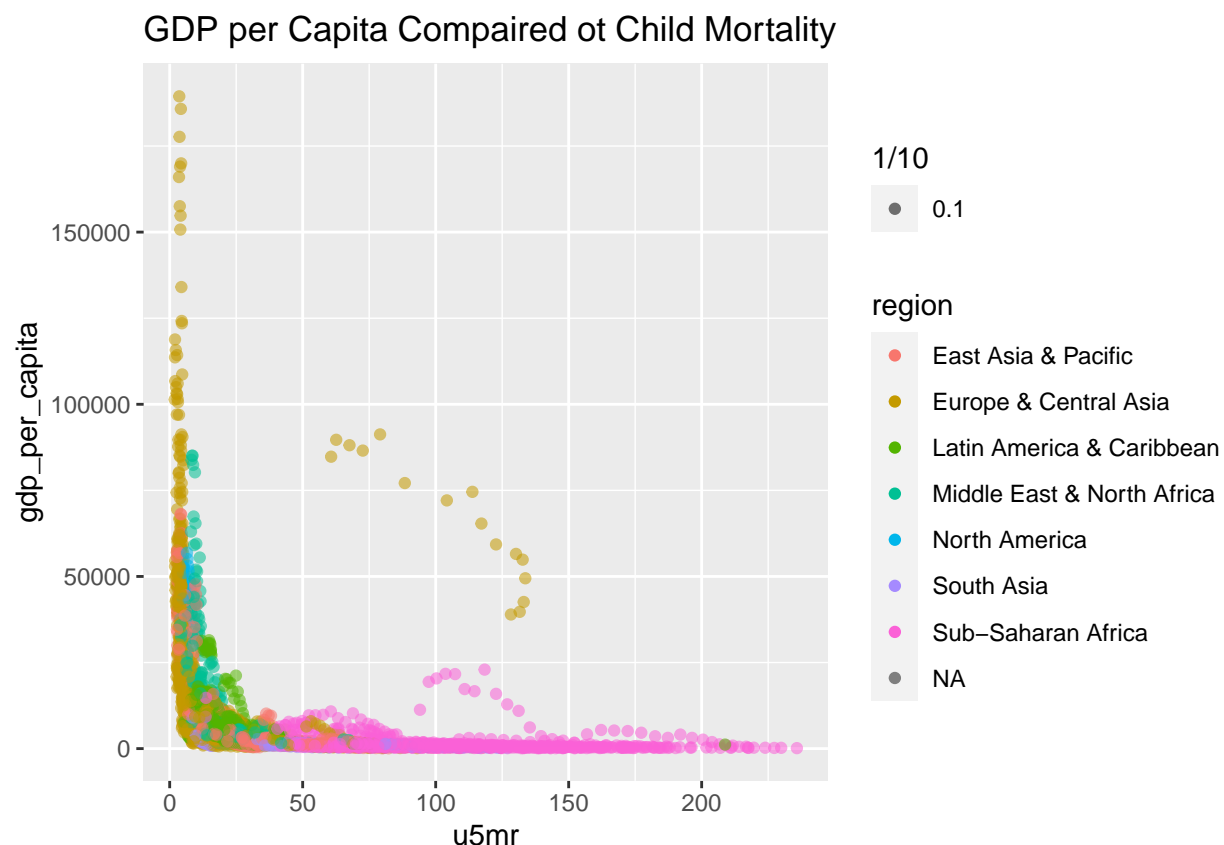
Warning: Removed 329 rows containing missing values (geom_point).

GDP Growth Compared to Child Mortality



```
ggplot(data = comb_df) +
  geom_point(mapping = aes(x=u5mr, y=gdp_per_capita, color=region, alpha=1/10)) +
  labs(
    title = "GDP per Capita Compared to Child Mortality"
  )
```

Warning: Removed 334 rows containing missing values (geom_point).



When we graph gdp grow agains child mortality we can see that even at the same growth rate Sub-Saharan African countries have higher child mortality rates.

```
t <- comb_df %>%
  select(gdp, gdp_growth, gdp_per_capita, u5mr, foreign_direct_investment_net, population_total, population_growth,
         drop_na())

comb_df.rcorr = rcorr(as.matrix(t))
comb_df.rcorr
```

```
##          gdp gdp_growth gdp_per_capita  u5mr
## gdp          1.00    -0.08         0.29 -0.15
## gdp_growth   -0.08     1.00        -0.25  0.24
## gdp_per_capita 0.29    -0.25         1.00 -0.42
## u5mr          -0.15     0.24        -0.42  1.00
## foreign_direct_investment_net 0.21    -0.11         0.20 -0.03
## population_total 0.35     0.12        -0.06  0.06
## population_growth -0.09     0.18        -0.14  0.52
## co2_emissions   0.32    -0.18         0.63 -0.52
## cell_subs       0.11    -0.27         0.50 -0.56
## consumer_price_index -0.09     0.08        -0.28  0.22
## fertility_rate   -0.15     0.21        -0.44  0.85
## enrollment_primary_secondary 0.07    -0.12         0.24 -0.69
## enrollment_secondary_school 0.15    -0.27         0.57 -0.81
## foreign_direct_investment_net population_total
```

## gdp	0.21	0.35
## gdp_growth	-0.11	0.12
## gdp_per_capita	0.20	-0.06
## u5mr	-0.03	0.06
## foreign_direct_investment_net	1.00	-0.19
## population_total	-0.19	1.00
## population_growth	-0.02	0.00
## co2_emissions	0.11	-0.02
## cell_subs	0.06	-0.08
## consumer_price_index	-0.05	0.03
## fertility_rate	-0.04	-0.02
## enrollment_primary_secondary	0.00	-0.04
## enrollment_secondary_school	0.06	-0.08
##	population_growth	co2_emissions
## gdp	-0.09	0.32
## gdp_growth	0.18	-0.18
## gdp_per_capita	-0.14	0.63
## u5mr	0.52	-0.52
## foreign_direct_investment_net	-0.02	0.11
## population_total	0.00	-0.02
## population_growth	1.00	-0.07
## co2_emissions	-0.07	1.00
## cell_subs	-0.29	0.46
## consumer_price_index	0.10	-0.22
## fertility_rate	0.71	-0.50
## enrollment_primary_secondary	-0.39	0.33
## enrollment_secondary_school	-0.55	0.58
##	consumer_price_index	fertility_rate
## gdp	-0.09	-0.15
## gdp_growth	0.08	0.21
## gdp_per_capita	-0.28	-0.44
## u5mr	0.22	0.85
## foreign_direct_investment_net	-0.05	-0.04
## population_total	0.03	-0.02
## population_growth	0.10	0.71
## co2_emissions	-0.22	-0.50
## cell_subs	-0.25	-0.53
## consumer_price_index	1.00	0.21
## fertility_rate	0.21	1.00
## enrollment_primary_secondary	-0.14	-0.65
## enrollment_secondary_school	-0.27	-0.82
##	enrollment_primary_secondary	
## gdp	0.07	
## gdp_growth	-0.12	
## gdp_per_capita	0.24	
## u5mr	-0.69	
## foreign_direct_investment_net	0.00	
## population_total	-0.04	
## population_growth	-0.39	
## co2_emissions	0.33	
## cell_subs	0.37	
## consumer_price_index	-0.14	
## fertility_rate	-0.65	
## enrollment_primary_secondary	1.00	

```

## enrollment_secondary_school          0.61
## enrollment_secondary_school
## gdp                                  0.15
## gdp_growth                         -0.27
## gdp_per_capita                      0.57
## u5mr                               -0.81
## foreign_direct_investment_net       0.06
## population_total                   -0.08
## population_growth                   -0.55
## co2_emissions                      0.58
## cell_subs                          0.61
## consumer_price_index                -0.27
## fertility_rate                      -0.82
## enrollment_primary_secondary         0.61
## enrollment_secondary_school         1.00
##
## n= 1715
##
## P
##
##          gdp      gdp_growth gdp_per_capita u5mr
## gdp          0.0015      0.0000      0.0000
## gdp_growth   0.0015      0.0000      0.0000
## gdp_per_capita 0.0000 0.0000      0.0000
## u5mr          0.0000 0.0000      0.0000
## foreign_direct_investment_net 0.0000 0.0000      0.0000      0.1605
## population_total 0.0000 0.0000      0.0090      0.0173
## population_growth 0.0001 0.0000      0.0000      0.0000
## co2_emissions  0.0000 0.0000      0.0000      0.0000
## cell_subs      0.0000 0.0000      0.0000      0.0000
## consumer_price_index 0.0002 0.0012      0.0000      0.0000
## fertility_rate  0.0000 0.0000      0.0000      0.0000
## enrollment_primary_secondary 0.0070 0.0000      0.0000      0.0000
## enrollment_secondary_school 0.0000 0.0000      0.0000      0.0000
##
##          foreign_direct_investment_net population_total
## gdp          0.0000      0.0000
## gdp_growth   0.0000      0.0000
## gdp_per_capita 0.0000      0.0090
## u5mr          0.1605      0.0173
## foreign_direct_investment_net 0.0000
## population_total 0.0000
## population_growth 0.4273      0.8773
## co2_emissions  0.0000      0.3913
## cell_subs      0.0074      0.0005
## consumer_price_index 0.0353      0.1532
## fertility_rate  0.0942      0.4802
## enrollment_primary_secondary 0.8668      0.1343
## enrollment_secondary_school 0.0125      0.0016
##
##          population_growth co2_emissions cell_subs
## gdp          0.0001      0.0000      0.0000
## gdp_growth   0.0000      0.0000      0.0000
## gdp_per_capita 0.0000      0.0000      0.0000
## u5mr          0.0000      0.0000      0.0000
## foreign_direct_investment_net 0.4273      0.0000      0.0074

```

```

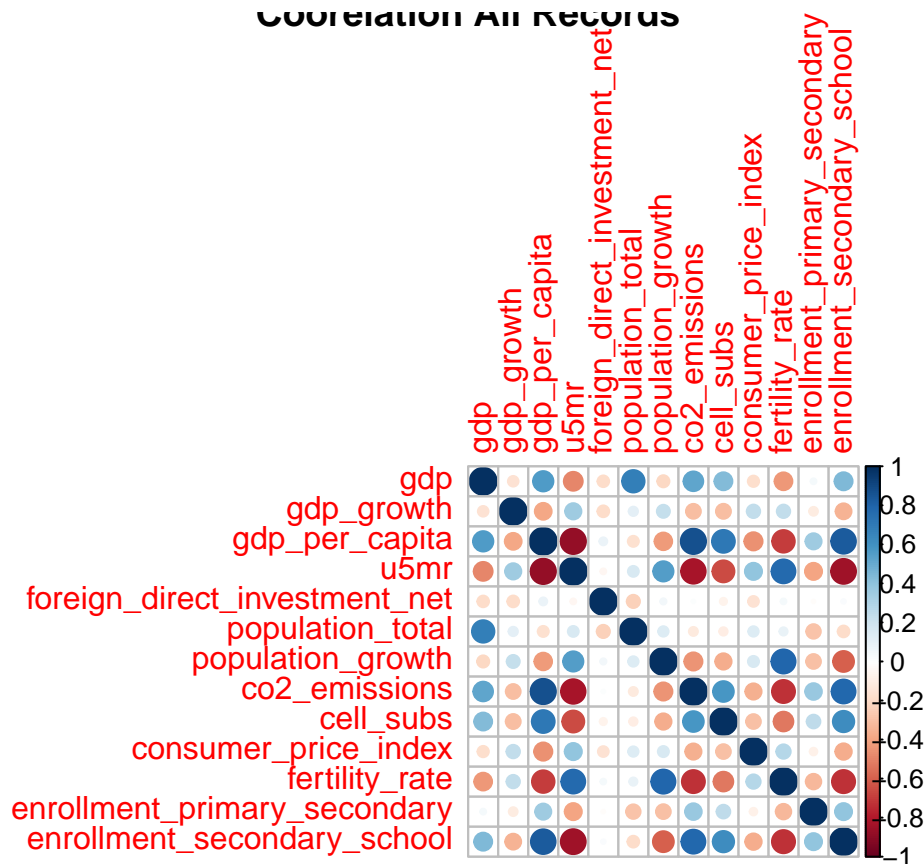
## population_total      0.8773      0.3913      0.0005
## population_growth      0.0024      0.0000
## co2_emissions          0.0024      0.0000
## cell_subs              0.0000      0.0000
## consumer_price_index    0.0000      0.0000      0.0000
## fertility_rate          0.0000      0.0000      0.0000
## enrollment_primary_secondary 0.0000      0.0000      0.0000
## enrollment_secondary_school 0.0000      0.0000      0.0000
##
## consumer_price_index fertility_rate
## gdp                    0.0002      0.0000
## gdp_growth              0.0012      0.0000
## gdp_per_capita          0.0000      0.0000
## u5mr                    0.0000      0.0000
## foreign_direct_investment_net 0.0353      0.0942
## population_total        0.1532      0.4802
## population_growth        0.0000      0.0000
## co2_emissions            0.0000      0.0000
## cell_subs                0.0000      0.0000
## consumer_price_index      0.0000
## fertility_rate            0.0000
## enrollment_primary_secondary 0.0000      0.0000
## enrollment_secondary_school 0.0000      0.0000
##
## enrollment_primary_secondary
## gdp                    0.0070
## gdp_growth              0.0000
## gdp_per_capita          0.0000
## u5mr                    0.0000
## foreign_direct_investment_net 0.8668
## population_total        0.1343
## population_growth        0.0000
## co2_emissions            0.0000
## cell_subs                0.0000
## consumer_price_index      0.0000
## fertility_rate            0.0000
## enrollment_primary_secondary
## enrollment_secondary_school 0.0000
##
## enrollment_secondary_school
## gdp                    0.0000
## gdp_growth              0.0000
## gdp_per_capita          0.0000
## u5mr                    0.0000
## foreign_direct_investment_net 0.0125
## population_total        0.0016
## population_growth        0.0000
## co2_emissions            0.0000
## cell_subs                0.0000
## consumer_price_index      0.0000
## fertility_rate            0.0000
## enrollment_primary_secondary 0.0000
## enrollment_secondary_school

```

```

comb_cor = cor(t, method = c("spearman"))
corrplot(comb_cor, title="Coorelation All Records")

```



By surveying a broad set of country variable we can see some interesting correlation. We can see relatively strong negative correlations between child mortality and gdp per capita, co2 emission, cell phone subscriptions and enrollment in secondary school. We can also see strong positive correlations between child mortality and fertility rates.

```
t <- comb_df %>%
  filter(region_id == "SSF") %>%
  select(gdp, gdp_growth, gdp_per_capita, u5mr, foreign_direct_investment_net, population_total, population_growth, co2_emissions, cell_subs, consumer_price_index, fertility_rate, enrollment_primary_secondary_school)
drop_na()

comb_df.rcorr = rcorr(as.matrix(t))
comb_df.rcorr
```

```
##           gdp gdp_growth gdp_per_capita u5mr
## gdp          1.00      0.03          0.20 0.03
## gdp_growth    0.03      1.00         -0.09 0.04
## gdp_per_capita 0.20     -0.09          1.00 -0.61
## u5mr           0.03      0.04         -0.61 1.00
## foreign_direct_investment_net -0.38    -0.08         -0.16 0.11
## population_total 0.72      0.25         -0.13 0.23
## population_growth -0.06     0.16         -0.56 0.51
## co2_emissions   0.39     -0.16          0.80 -0.50
## cell_subs       0.26     -0.04          0.61 -0.58
## consumer_price_index 0.03    -0.06         -0.10 -0.04
## fertility_rate  -0.10     0.14         -0.74 0.79
```

## enrollment_primary_secondary	0.07	-0.02	0.45	-0.73
## enrollment_secondary_school	0.29	-0.09	0.75	-0.72
##	foreign_direct_investment_net population_total			
## gdp			-0.38	0.72
## gdp_growth			-0.08	0.25
## gdp_per_capita			-0.16	-0.13
## u5mr			0.11	0.23
## foreign_direct_investment_net			1.00	-0.37
## population_total			-0.37	1.00
## population_growth			0.12	0.20
## co2_emissions			-0.14	-0.04
## cell_subs			-0.21	-0.06
## consumer_price_index			-0.01	0.19
## fertility_rate			0.14	0.24
## enrollment_primary_secondary			-0.10	-0.17
## enrollment_secondary_school			-0.21	-0.08
##	population_growth co2_emissions cell_subs			
## gdp		-0.06	0.39	0.26
## gdp_growth		0.16	-0.16	-0.04
## gdp_per_capita		-0.56	0.80	0.61
## u5mr		0.51	-0.50	-0.58
## foreign_direct_investment_net		0.12	-0.14	-0.21
## population_total		0.20	-0.04	-0.06
## population_growth		1.00	-0.55	-0.33
## co2_emissions		-0.55	1.00	0.45
## cell_subs		-0.33	0.45	1.00
## consumer_price_index		0.00	-0.10	-0.13
## fertility_rate		0.83	-0.71	-0.52
## enrollment_primary_secondary		-0.58	0.43	0.48
## enrollment_secondary_school		-0.65	0.74	0.62
##	consumer_price_index fertility_rate			
## gdp		0.03	-0.10	
## gdp_growth		-0.06	0.14	
## gdp_per_capita		-0.10	-0.74	
## u5mr		-0.04	0.79	
## foreign_direct_investment_net		-0.01	0.14	
## population_total		0.19	0.24	
## population_growth		0.00	0.83	
## co2_emissions		-0.10	-0.71	
## cell_subs		-0.13	-0.52	
## consumer_price_index		1.00	0.00	
## fertility_rate		0.00	1.00	
## enrollment_primary_secondary		0.04	-0.73	
## enrollment_secondary_school		-0.06	-0.87	
##	enrollment_primary_secondary			
## gdp		0.07		
## gdp_growth		-0.02		
## gdp_per_capita		0.45		
## u5mr		-0.73		
## foreign_direct_investment_net		-0.10		
## population_total		-0.17		
## population_growth		-0.58		
## co2_emissions		0.43		
## cell_subs		0.48		

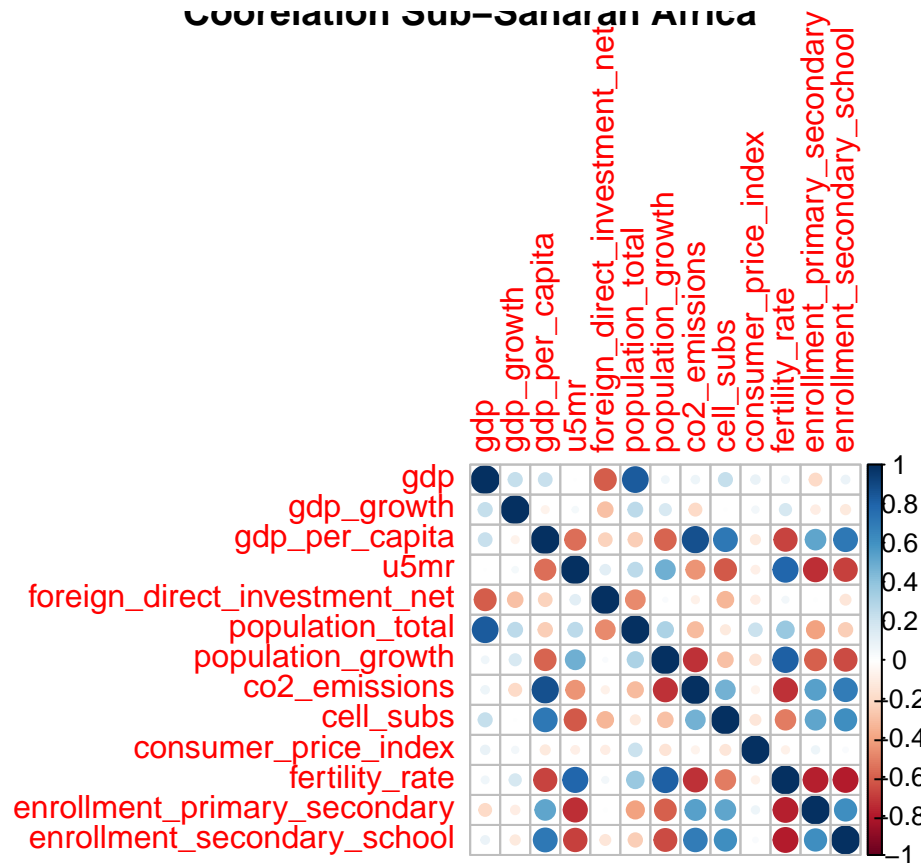
```

## consumer_price_index          0.04
## fertility_rate                 -0.73
## enrollment_primary_secondary   1.00
## enrollment_secondary_school    0.61
##                               enrollment_secondary_school
## gdp                           0.29
## gdp_growth                    -0.09
## gdp_per_capita                 0.75
## u5mr                          -0.72
## foreign_direct_investment_net -0.21
## population_total              -0.08
## population_growth             -0.65
## co2_emissions                 0.74
## cell_subs                     0.62
## consumer_price_index          -0.06
## fertility_rate                 -0.87
## enrollment_primary_secondary   0.61
## enrollment_secondary_school    1.00
##
## n= 335
##
##
## P
##                               gdp      gdp_growth gdp_per_capita u5mr
## gdp                          0.5496      0.0003      0.6205
## gdp_growth                   0.5496      0.1157      0.4236
## gdp_per_capita               0.0003 0.1157      0.0000
## u5mr                         0.6205 0.4236      0.0000
## foreign_direct_investment_net 0.0000 0.1635      0.0034      0.0376
## population_total             0.0000 0.0000      0.0214      0.0000
## population_growth            0.3066 0.0027      0.0000      0.0000
## co2_emissions                0.0000 0.0027      0.0000      0.0000
## cell_subs                    0.0000 0.5119      0.0000      0.0000
## consumer_price_index         0.5590 0.2766      0.0746      0.4843
## fertility_rate                0.0704 0.0129      0.0000      0.0000
## enrollment_primary_secondary  0.1866 0.6955      0.0000      0.0000
## enrollment_secondary_school   0.0000 0.0950      0.0000      0.0000
##                               foreign_direct_investment_net population_total
## gdp                          0.0000      0.0000
## gdp_growth                   0.1635      0.0000
## gdp_per_capita               0.0034      0.0214
## u5mr                         0.0376      0.0000
## foreign_direct_investment_net 0.0000
## population_total             0.0000
## population_growth            0.0292      0.0002
## co2_emissions                0.0107      0.4995
## cell_subs                    0.0000      0.2470
## consumer_price_index         0.8197      0.0005
## fertility_rate                0.0099      0.0000
## enrollment_primary_secondary  0.0581      0.0023
## enrollment_secondary_school   0.0001      0.1384
##                               population_growth co2_emissions cell_subs
## gdp                          0.3066      0.0000      0.0000
## gdp_growth                   0.0027      0.0027      0.5119

```


## gdp_per_capita	0.0000	0.0000	0.0000
## u5mr	0.0000	0.0000	0.0000
## foreign_direct_investment_net	0.0292	0.0107	0.0000
## population_total	0.0002	0.4995	0.2470
## population_growth		0.0000	0.0000
## co2_emissions	0.0000		0.0000
## cell_subs	0.0000	0.0000	
## consumer_price_index	0.9849	0.0808	0.0189
## fertility_rate	0.0000	0.0000	0.0000
## enrollment_primary_secondary	0.0000	0.0000	0.0000
## enrollment_secondary_school	0.0000	0.0000	0.0000
##	consumer_price_index	fertility_rate	
## gdp	0.5590	0.0704	
## gdp_growth	0.2766	0.0129	
## gdp_per_capita	0.0746	0.0000	
## u5mr	0.4843	0.0000	
## foreign_direct_investment_net	0.8197	0.0099	
## population_total	0.0005	0.0000	
## population_growth	0.9849	0.0000	
## co2_emissions	0.0808	0.0000	
## cell_subs	0.0189	0.0000	
## consumer_price_index		0.9556	
## fertility_rate	0.9556		
## enrollment_primary_secondary	0.4945	0.0000	
## enrollment_secondary_school	0.2910	0.0000	
##	enrollment_primary_secondary		
## gdp	0.1866		
## gdp_growth	0.6955		
## gdp_per_capita	0.0000		
## u5mr	0.0000		
## foreign_direct_investment_net	0.0581		
## population_total	0.0023		
## population_growth	0.0000		
## co2_emissions	0.0000		
## cell_subs	0.0000		
## consumer_price_index	0.4945		
## fertility_rate	0.0000		
## enrollment_primary_secondary			
## enrollment_secondary_school	0.0000		
##	enrollment_secondary_school		
## gdp	0.0000		
## gdp_growth	0.0950		
## gdp_per_capita	0.0000		
## u5mr	0.0000		
## foreign_direct_investment_net	0.0001		
## population_total	0.1384		
## population_growth	0.0000		
## co2_emissions	0.0000		
## cell_subs	0.0000		
## consumer_price_index	0.2910		
## fertility_rate	0.0000		
## enrollment_primary_secondary	0.0000		
## enrollment_secondary_school			

```
comb_cor = cor(t, method = c("spearman"))
corrplot(comb_cor, title="Coorelation Sub-Saharan Africa")
```



By surveying the Sub-Saharan Africa country variable we can see some interesting correlation. We can see relatively strong negative correlations between enrollment in primary school and enrollment in secondary school. We can also see strong positive correlations between child mortality and fertility rates.

There are also some expected observation negative correlation fertility rate and enrollment in primary and secondary school positive correlation gdp per capita and enrollment in primary and secondary school positive correlation between co2 emissions and gdp per capitat

```
t <- comb_df %>%
  filter(region_id == "NAC") %>%
  select(gdp, gdp_growth, gdp_per_capita, u5mr, foreign_direct_investment_net, population_total, population_growth, co2_emissions, cell_subs, consumer_price_index, fertility_rate, enrollment_primary_secondary, enrollment_secondary_school)
  drop_na()

comb_df.rcorr = rcorr(as.matrix(t))
comb_df.rcorr
```

```
##          gdp gdp_growth gdp_per_capita u5mr
## gdp      1.00   -0.19      0.34   0.74
## gdp_growth -0.19    1.00     -0.21 -0.04
## gdp_per_capita 0.34   -0.21    1.00 -0.25
## u5mr        0.74   -0.04     -0.25  1.00
## foreign_direct_investment_net 0.29   -0.34    0.16  0.17
```

## population_total	0.98	-0.18	0.18	0.86
## population_growth	-0.65	0.00	-0.31	-0.21
## co2_emissions	0.34	0.17	-0.49	0.85
## cell_subs	0.51	-0.32	0.90	-0.16
## consumer_price_index	0.08	0.41	-0.28	0.41
## fertility_rate	0.82	-0.21	0.04	0.93
## enrollment_primary_secondary	0.39	-0.20	0.03	0.28
## enrollment_secondary_school	-0.81	0.03	0.13	-0.93
##	foreign_direct_investment_net population_total			
## gdp			0.29	0.98
## gdp_growth			-0.34	-0.18
## gdp_per_capita			0.16	0.18
## u5mr			0.17	0.86
## foreign_direct_investment_net			1.00	0.27
## population_total			0.27	1.00
## population_growth			-0.36	-0.56
## co2_emissions			-0.05	0.52
## cell_subs			0.26	0.33
## consumer_price_index			-0.05	0.17
## fertility_rate			0.22	0.91
## enrollment_primary_secondary			0.14	0.42
## enrollment_secondary_school			-0.21	-0.87
##	population_growth co2_emissions cell_subs			
## gdp		-0.65	0.34	0.51
## gdp_growth		0.00	0.17	-0.32
## gdp_per_capita		-0.31	-0.49	0.90
## u5mr		-0.21	0.85	-0.16
## foreign_direct_investment_net		-0.36	-0.05	0.26
## population_total		-0.56	0.52	0.33
## population_growth		1.00	0.15	-0.57
## co2_emissions		0.15	1.00	-0.53
## cell_subs		-0.57	-0.53	1.00
## consumer_price_index		0.14	0.58	-0.39
## fertility_rate		-0.23	0.75	0.10
## enrollment_primary_secondary		-0.34	0.21	0.14
## enrollment_secondary_school		0.33	-0.66	-0.01
##	consumer_price_index fertility_rate			
## gdp		0.08		0.82
## gdp_growth		0.41		-0.21
## gdp_per_capita		-0.28		0.04
## u5mr		0.41		0.93
## foreign_direct_investment_net		-0.05		0.22
## population_total		0.17		0.91
## population_growth		0.14		-0.23
## co2_emissions		0.58		0.75
## cell_subs		-0.39		0.10
## consumer_price_index		1.00		0.33
## fertility_rate		0.33		1.00
## enrollment_primary_secondary		-0.05		0.34
## enrollment_secondary_school		-0.33		-0.86
##	enrollment_primary_secondary			
## gdp			0.39	
## gdp_growth			-0.20	
## gdp_per_capita			0.03	

```

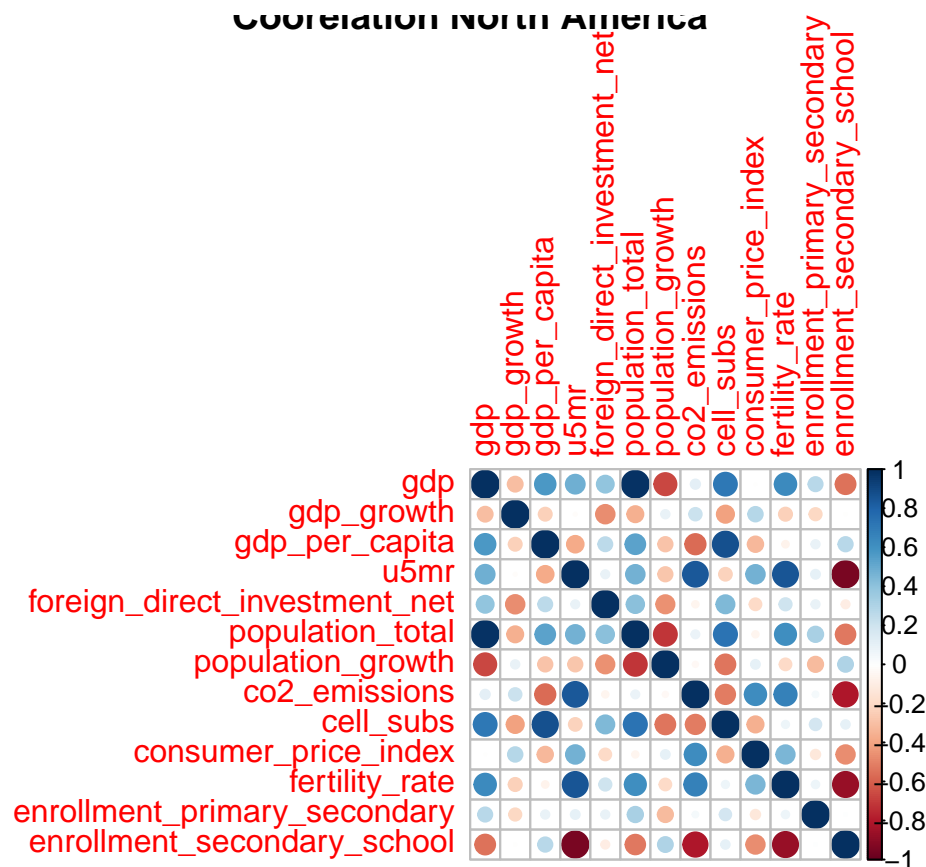
## u5mr 0.28
## foreign_direct_investment_net 0.14
## population_total 0.42
## population_growth -0.34
## co2_emissions 0.21
## cell_subs 0.14
## consumer_price_index -0.05
## fertility_rate 0.34
## enrollment_primary_secondary 1.00
## enrollment_secondary_school -0.09
## enrollment_secondary_school
## gdp -0.81
## gdp_growth 0.03
## gdp_per_capita 0.13
## u5mr -0.93
## foreign_direct_investment_net -0.21
## population_total -0.87
## population_growth 0.33
## co2_emissions -0.66
## cell_subs -0.01
## consumer_price_index -0.33
## fertility_rate -0.86
## enrollment_primary_secondary -0.09
## enrollment_secondary_school 1.00
##
## n= 28
##
##
## P
##
## gdp gdp_growth gdp_per_capita u5mr
## gdp 0.3307 0.0779 0.0000
## gdp_growth 0.3307 0.2790 0.8256
## gdp_per_capita 0.0779 0.2790 0.1968
## u5mr 0.0000 0.8256 0.1968
## foreign_direct_investment_net 0.1307 0.0786 0.4062 0.3895
## population_total 0.0000 0.3636 0.3497 0.0000
## population_growth 0.0002 0.9962 0.1087 0.2823
## co2_emissions 0.0763 0.3790 0.0087 0.0000
## cell_subs 0.0055 0.0957 0.0000 0.4040
## consumer_price_index 0.6688 0.0307 0.1471 0.0310
## fertility_rate 0.0000 0.2875 0.8410 0.0000
## enrollment_primary_secondary 0.0381 0.3196 0.8828 0.1453
## enrollment_secondary_school 0.0000 0.8612 0.4987 0.0000
## foreign_direct_investment_net population_total
## gdp 0.1307 0.0000
## gdp_growth 0.0786 0.3636
## gdp_per_capita 0.4062 0.3497
## u5mr 0.3895 0.0000
## foreign_direct_investment_net 0.1579
## population_total 0.1579
## population_growth 0.0598 0.0019
## co2_emissions 0.8073 0.0047
## cell_subs 0.1831 0.0830
## consumer_price_index 0.8186 0.3856

```

## fertility_rate	0.2646		0.0000
## enrollment_primary_secondary	0.4700		0.0254
## enrollment_secondary_school	0.2902		0.0000
##	population_growth	co2_emissions	cell_subs
## gdp	0.0002	0.0763	0.0055
## gdp_growth	0.9962	0.3790	0.0957
## gdp_per_capita	0.1087	0.0087	0.0000
## u5mr	0.2823	0.0000	0.4040
## foreign_direct_investment_net	0.0598	0.8073	0.1831
## population_total	0.0019	0.0047	0.0830
## population_growth		0.4444	0.0016
## co2_emissions	0.4444		0.0035
## cell_subs	0.0016	0.0035	
## consumer_price_index	0.4889	0.0011	0.0419
## fertility_rate	0.2319	0.0000	0.6265
## enrollment_primary_secondary	0.0762	0.2845	0.4634
## enrollment_secondary_school	0.0823	0.0001	0.9658
##	consumer_price_index	fertility_rate	
## gdp	0.6688	0.0000	
## gdp_growth	0.0307	0.2875	
## gdp_per_capita	0.1471	0.8410	
## u5mr	0.0310	0.0000	
## foreign_direct_investment_net	0.8186	0.2646	
## population_total	0.3856	0.0000	
## population_growth	0.4889	0.2319	
## co2_emissions	0.0011	0.0000	
## cell_subs	0.0419	0.6265	
## consumer_price_index		0.0909	
## fertility_rate	0.0909		
## enrollment_primary_secondary	0.7989	0.0788	
## enrollment_secondary_school	0.0864	0.0000	
##	enrollment_primary_secondary		
## gdp	0.0381		
## gdp_growth	0.3196		
## gdp_per_capita	0.8828		
## u5mr	0.1453		
## foreign_direct_investment_net	0.4700		
## population_total	0.0254		
## population_growth	0.0762		
## co2_emissions	0.2845		
## cell_subs	0.4634		
## consumer_price_index	0.7989		
## fertility_rate	0.0788		
## enrollment_primary_secondary			
## enrollment_secondary_school	0.6570		
##	enrollment_secondary_school		
## gdp	0.0000		
## gdp_growth	0.8612		
## gdp_per_capita	0.4987		
## u5mr	0.0000		
## foreign_direct_investment_net	0.2902		
## population_total	0.0000		
## population_growth	0.0823		
## co2_emissions	0.0001		

```
## cell_subs 0.9658
## consumer_price_index 0.0864
## fertility_rate 0.0000
## enrollment_primary_secondary 0.6570
## enrollment_secondary_school
```

```
comb_cor = cor(t, method = c("spearman"))
corrplot(comb_cor, title="Coorelation North America")
```



When you survey the North American country variable we can see some interesting correlation. We can see relatively strong negative correlations between child mortality and enrollment in secondary school. We can also see strong positive correlations between child mortality and fertility rates.

```
t <- comb_df %>%
  filter(income_group == "Low income") %>%
  select(gdp, gdp_growth, gdp_per_capita, u5mr, foreign_direct_investment_net, population_total, populat
  drop_na()

comb_df.rcorr = rcorr(as.matrix(t))
comb_df.rcorr
```

```
##          gdp gdp_growth gdp_per_capita  u5mr
## gdp      1.00      0.15      0.67 -0.43
## gdp_growth 0.15      1.00     -0.05 -0.04
## gdp_per_capita 0.67     -0.05      1.00 -0.57
```

## u5mr	-0.43	-0.04	-0.57	1.00
## foreign_direct_investment_net	-0.41	-0.17	-0.20	0.17
## population_total	0.59	0.32	-0.05	-0.16
## population_growth	-0.22	0.07	-0.28	0.36
## co2_emissions	0.34	-0.05	0.70	-0.53
## cell_subs	0.27	-0.03	0.47	-0.35
## consumer_price_index	0.29	-0.15	0.11	-0.16
## fertility_rate	-0.37	0.05	-0.57	0.82
## enrollment_primary_secondary	0.06	0.07	0.18	-0.55
## enrollment_secondary_school	0.43	0.03	0.58	-0.65
##	foreign_direct_investment_net population_total			
## gdp			-0.41	0.59
## gdp_growth			-0.17	0.32
## gdp_per_capita			-0.20	-0.05
## u5mr			0.17	-0.16
## foreign_direct_investment_net			1.00	-0.28
## population_total			-0.28	1.00
## population_growth			0.07	0.02
## co2_emissions			0.02	-0.07
## cell_subs			-0.21	-0.07
## consumer_price_index			-0.05	0.25
## fertility_rate			0.13	-0.03
## enrollment_primary_secondary			-0.14	-0.09
## enrollment_secondary_school			-0.11	0.14
##	population_growth co2_emissions cell_subs			
## gdp		-0.22	0.34	0.27
## gdp_growth		0.07	-0.05	-0.03
## gdp_per_capita		-0.28	0.70	0.47
## u5mr		0.36	-0.53	-0.35
## foreign_direct_investment_net		0.07	0.02	-0.21
## population_total		0.02	-0.07	-0.07
## population_growth		1.00	-0.14	-0.11
## co2_emissions		-0.14	1.00	-0.03
## cell_subs		-0.11	-0.03	1.00
## consumer_price_index		-0.38	-0.06	-0.03
## fertility_rate		0.67	-0.49	-0.30
## enrollment_primary_secondary		-0.26	0.11	0.25
## enrollment_secondary_school		-0.33	0.57	0.40
##	consumer_price_index fertility_rate			
## gdp		0.29	-0.37	
## gdp_growth		-0.15	0.05	
## gdp_per_capita		0.11	-0.57	
## u5mr		-0.16	0.82	
## foreign_direct_investment_net		-0.05	0.13	
## population_total		0.25	-0.03	
## population_growth		-0.38	0.67	
## co2_emissions		-0.06	-0.49	
## cell_subs		-0.03	-0.30	
## consumer_price_index		1.00	-0.24	
## fertility_rate		-0.24	1.00	
## enrollment_primary_secondary		0.09	-0.52	
## enrollment_secondary_school		0.06	-0.68	
##	enrollment_primary_secondary			
## gdp			0.06	

```

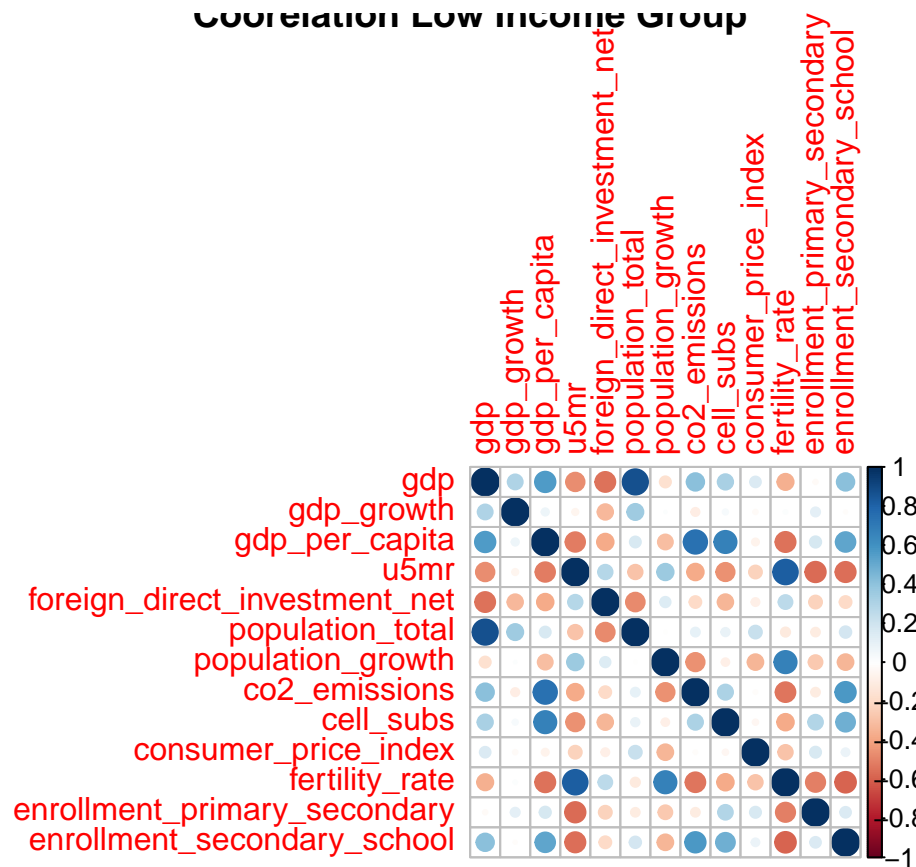
## gdp_growth 0.07
## gdp_per_capita 0.18
## u5mr -0.55
## foreign_direct_investment_net -0.14
## population_total -0.09
## population_growth -0.26
## co2_emissions 0.11
## cell_subs 0.25
## consumer_price_index 0.09
## fertility_rate -0.52
## enrollment_primary_secondary 1.00
## enrollment_secondary_school 0.22
## enrollment_secondary_school
## gdp 0.43
## gdp_growth 0.03
## gdp_per_capita 0.58
## u5mr -0.65
## foreign_direct_investment_net -0.11
## population_total 0.14
## population_growth -0.33
## co2_emissions 0.57
## cell_subs 0.40
## consumer_price_index 0.06
## fertility_rate -0.68
## enrollment_primary_secondary 0.22
## enrollment_secondary_school 1.00
##
## n= 174
##
##
## P
##
## gdp gdp_growth gdp_per_capita u5mr
## gdp 0.0556 0.0000 0.0000
## gdp_growth 0.0556 0.5473 0.6009
## gdp_per_capita 0.0000 0.5473 0.0000
## u5mr 0.0000 0.6009 0.0000
## foreign_direct_investment_net 0.0000 0.0265 0.0089 0.0236
## population_total 0.0000 0.0000 0.5219 0.0406
## population_growth 0.0039 0.3819 0.0002 0.0000
## co2_emissions 0.0000 0.5050 0.0000 0.0000
## cell_subs 0.0004 0.6627 0.0000 0.0000
## consumer_price_index 0.0001 0.0513 0.1381 0.0369
## fertility_rate 0.0000 0.5375 0.0000 0.0000
## enrollment_primary_secondary 0.4403 0.3761 0.0191 0.0000
## enrollment_secondary_school 0.0000 0.6585 0.0000 0.0000
## foreign_direct_investment_net population_total
## gdp 0.0000 0.0000
## gdp_growth 0.0265 0.0000
## gdp_per_capita 0.0089 0.5219
## u5mr 0.0236 0.0406
## foreign_direct_investment_net 0.0002
## population_total 0.0002
## population_growth 0.3878 0.8248
## co2_emissions 0.8192 0.3310

```


## cell_subs	0.0052		0.3761
## consumer_price_index	0.5309		0.0010
## fertility_rate	0.0778		0.6847
## enrollment_primary_secondary	0.0717		0.2473
## enrollment_secondary_school	0.1438		0.0654
##	population_growth	co2_emissions	cell_subs
## gdp	0.0039	0.0000	0.0004
## gdp_growth	0.3819	0.5050	0.6627
## gdp_per_capita	0.0002	0.0000	0.0000
## u5mr	0.0000	0.0000	0.0000
## foreign_direct_investment_net	0.3878	0.8192	0.0052
## population_total	0.8248	0.3310	0.3761
## population_growth		0.0669	0.1337
## co2_emissions	0.0669		0.6716
## cell_subs	0.1337	0.6716	
## consumer_price_index	0.0000	0.4530	0.6797
## fertility_rate	0.0000	0.0000	0.0000
## enrollment_primary_secondary	0.0005	0.1467	0.0008
## enrollment_secondary_school	0.0000	0.0000	0.0000
##	consumer_price_index	fertility_rate	
## gdp	0.0001	0.0000	
## gdp_growth	0.0513	0.5375	
## gdp_per_capita	0.1381	0.0000	
## u5mr	0.0369	0.0000	
## foreign_direct_investment_net	0.5309	0.0778	
## population_total	0.0010	0.6847	
## population_growth	0.0000	0.0000	
## co2_emissions	0.4530	0.0000	
## cell_subs	0.6797	0.0000	
## consumer_price_index		0.0012	
## fertility_rate	0.0012		
## enrollment_primary_secondary	0.2514	0.0000	
## enrollment_secondary_school	0.4270	0.0000	
##	enrollment_primary_secondary		
## gdp	0.4403		
## gdp_growth	0.3761		
## gdp_per_capita	0.0191		
## u5mr	0.0000		
## foreign_direct_investment_net	0.0717		
## population_total	0.2473		
## population_growth	0.0005		
## co2_emissions	0.1467		
## cell_subs	0.0008		
## consumer_price_index	0.2514		
## fertility_rate	0.0000		
## enrollment_primary_secondary			
## enrollment_secondary_school	0.0036		
##	enrollment_secondary_school		
## gdp	0.0000		
## gdp_growth	0.6585		
## gdp_per_capita	0.0000		
## u5mr	0.0000		
## foreign_direct_investment_net	0.1438		
## population_total	0.0654		

```
## population_growth      0.0000
## co2_emissions          0.0000
## cell_subs              0.0000
## consumer_price_index   0.4270
## fertility_rate         0.0000
## enrollment_primary_secondary 0.0036
## enrollment_secondary_school
```

```
comb_cor = cor(t, method = c("spearman"))
corrplot(comb_cor, title="Coorelation Low Income Group")
```



When you survey the low income country the correlations between specific variable are not as pronounced.

```
t <- comb_df %>%
  filter(income_group == "High income") %>%
  select(gdp, gdp_growth, gdp_per_capita, u5mr, foreign_direct_investment_net, population_total, population_growth, co2_emissions, cell_subs, consumer_price_index, fertility_rate, enrollment_primary_secondary, enrollment_secondary_school)
drop_na()

comb_df.rcorr = rcorr(as.matrix(t))
comb_df.rcorr
```

```
##          gdp gdp_growth gdp_per_capita  u5mr
## gdp      1.00   -0.06         0.19 -0.04
## gdp_growth -0.06    1.00        -0.11  0.02
## gdp_per_capita 0.19   -0.11         1.00  0.09
```

## u5mr	-0.04	0.02	0.09	1.00
## foreign_direct_investment_net	0.28	-0.05	0.15	0.06
## population_total	0.98	-0.05	0.12	-0.06
## population_growth	-0.02	0.10	0.22	0.05
## co2_emissions	0.24	0.09	0.25	-0.08
## cell_subs	-0.05	-0.18	0.34	-0.10
## consumer_price_index	-0.07	0.02	-0.23	-0.03
## fertility_rate	0.04	0.05	0.05	-0.02
## enrollment_primary_secondary	-0.07	0.09	-0.03	-0.15
## enrollment_secondary_school	-0.05	-0.09	0.26	-0.18
##	foreign_direct_investment_net population_total			
## gdp			0.28	0.98
## gdp_growth			-0.05	-0.05
## gdp_per_capita			0.15	0.12
## u5mr			0.06	-0.06
## foreign_direct_investment_net			1.00	0.26
## population_total			0.26	1.00
## population_growth			0.00	-0.05
## co2_emissions			0.03	0.22
## cell_subs			0.02	-0.09
## consumer_price_index			-0.05	-0.06
## fertility_rate			-0.02	-0.01
## enrollment_primary_secondary			-0.07	-0.09
## enrollment_secondary_school			-0.03	-0.07
##	population_growth co2_emissions cell_subs			
## gdp		-0.02	0.24	-0.05
## gdp_growth		0.10	0.09	-0.18
## gdp_per_capita		0.22	0.25	0.34
## u5mr		0.05	-0.08	-0.10
## foreign_direct_investment_net		0.00	0.03	0.02
## population_total		-0.05	0.22	-0.09
## population_growth		1.00	0.58	0.06
## co2_emissions		0.58	1.00	0.03
## cell_subs		0.06	0.03	1.00
## consumer_price_index		-0.03	-0.07	-0.10
## fertility_rate		0.60	0.33	-0.03
## enrollment_primary_secondary		0.19	0.14	-0.10
## enrollment_secondary_school		-0.10	-0.05	0.15
##	consumer_price_index fertility_rate			
## gdp		-0.07	0.04	
## gdp_growth		0.02	0.05	
## gdp_per_capita		-0.23	0.05	
## u5mr		-0.03	-0.02	
## foreign_direct_investment_net		-0.05	-0.02	
## population_total		-0.06	-0.01	
## population_growth		-0.03	0.60	
## co2_emissions		-0.07	0.33	
## cell_subs		-0.10	-0.03	
## consumer_price_index		1.00	0.05	
## fertility_rate		0.05	1.00	
## enrollment_primary_secondary		0.15	0.22	
## enrollment_secondary_school		-0.14	-0.03	
##	enrollment_primary_secondary			
## gdp			-0.07	

```

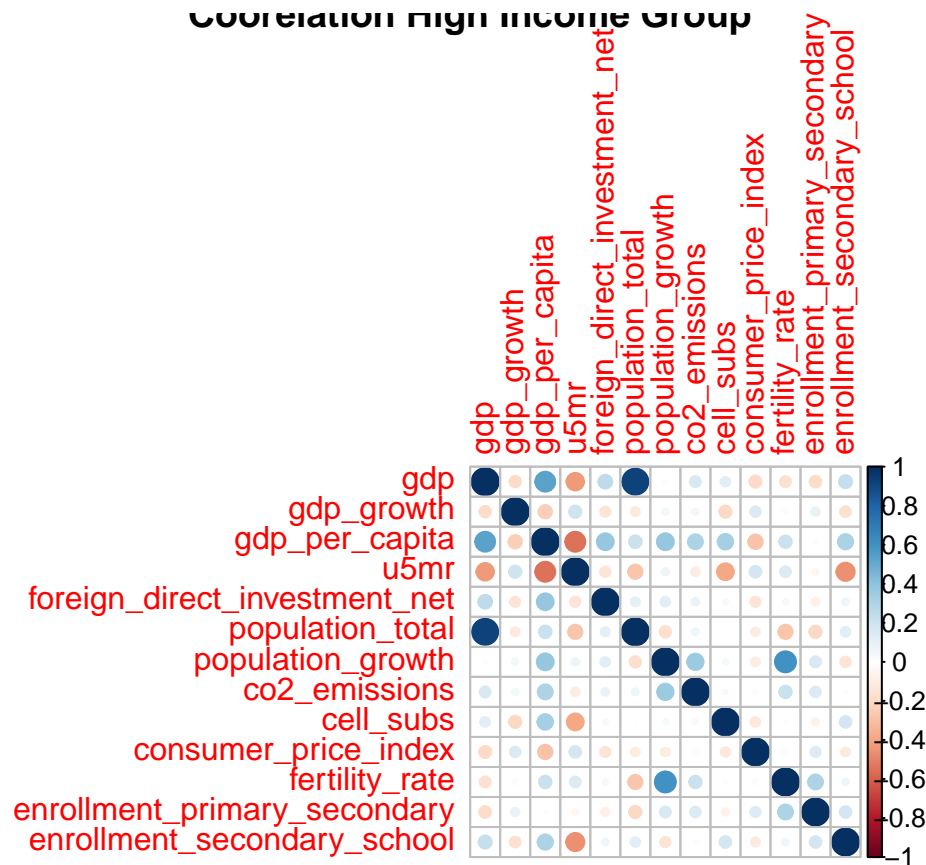
## gdp_growth 0.09
## gdp_per_capita -0.03
## u5mr -0.15
## foreign_direct_investment_net -0.07
## population_total -0.09
## population_growth 0.19
## co2_emissions 0.14
## cell_subs -0.10
## consumer_price_index 0.15
## fertility_rate 0.22
## enrollment_primary_secondary 1.00
## enrollment_secondary_school 0.37
## enrollment_secondary_school
## gdp -0.05
## gdp_growth -0.09
## gdp_per_capita 0.26
## u5mr -0.18
## foreign_direct_investment_net -0.03
## population_total -0.07
## population_growth -0.10
## co2_emissions -0.05
## cell_subs 0.15
## consumer_price_index -0.14
## fertility_rate -0.03
## enrollment_primary_secondary 0.37
## enrollment_secondary_school 1.00
##
## n= 666
##
##
## P
##
## gdp gdp_growth gdp_per_capita u5mr
## gdp 0.1503 0.0000 0.3144
## gdp_growth 0.1503 0.0056 0.6666
## gdp_per_capita 0.0000 0.0056 0.0176
## u5mr 0.3144 0.6666 0.0176
## foreign_direct_investment_net 0.0000 0.1820 0.0000 0.1441
## population_total 0.0000 0.2321 0.0017 0.1158
## population_growth 0.5222 0.0084 0.0000 0.2248
## co2_emissions 0.0000 0.0187 0.0000 0.0349
## cell_subs 0.2261 0.0000 0.0000 0.0138
## consumer_price_index 0.0599 0.5808 0.0000 0.4373
## fertility_rate 0.3525 0.2217 0.2057 0.5848
## enrollment_primary_secondary 0.0534 0.0223 0.4691 0.0000
## enrollment_secondary_school 0.1564 0.0179 0.0000 0.0000
## foreign_direct_investment_net population_total
## gdp 0.0000 0.0000
## gdp_growth 0.1820 0.2321
## gdp_per_capita 0.0000 0.0017
## u5mr 0.1441 0.1158
## foreign_direct_investment_net 0.0000
## population_total 0.0000
## population_growth 0.9541 0.1836
## co2_emissions 0.4254 0.0000

```

## cell_subs	0.6878		0.0216
## consumer_price_index	0.1756		0.0940
## fertility_rate	0.6865		0.7575
## enrollment_primary_secondary	0.0757		0.0147
## enrollment_secondary_school	0.4122		0.0690
##	population_growth	co2_emissions	cell_subs
## gdp	0.5222	0.0000	0.2261
## gdp_growth	0.0084	0.0187	0.0000
## gdp_per_capita	0.0000	0.0000	0.0000
## u5mr	0.2248	0.0349	0.0138
## foreign_direct_investment_net	0.9541	0.4254	0.6878
## population_total	0.1836	0.0000	0.0216
## population_growth		0.0000	0.0986
## co2_emissions	0.0000		0.4724
## cell_subs	0.0986	0.4724	
## consumer_price_index	0.4202	0.0837	0.0070
## fertility_rate	0.0000	0.0000	0.4491
## enrollment_primary_secondary	0.0000	0.0003	0.0117
## enrollment_secondary_school	0.0089	0.1607	0.0000
##	consumer_price_index	fertility_rate	
## gdp	0.0599	0.3525	
## gdp_growth	0.5808	0.2217	
## gdp_per_capita	0.0000	0.2057	
## u5mr	0.4373	0.5848	
## foreign_direct_investment_net	0.1756	0.6865	
## population_total	0.0940	0.7575	
## population_growth	0.4202	0.0000	
## co2_emissions	0.0837	0.0000	
## cell_subs	0.0070	0.4491	
## consumer_price_index		0.2415	
## fertility_rate	0.2415		
## enrollment_primary_secondary	0.0000	0.0000	
## enrollment_secondary_school	0.0003	0.3739	
##	enrollment_primary_secondary		
## gdp	0.0534		
## gdp_growth	0.0223		
## gdp_per_capita	0.4691		
## u5mr	0.0000		
## foreign_direct_investment_net	0.0757		
## population_total	0.0147		
## population_growth	0.0000		
## co2_emissions	0.0003		
## cell_subs	0.0117		
## consumer_price_index	0.0000		
## fertility_rate	0.0000		
## enrollment_primary_secondary			
## enrollment_secondary_school	0.0000		
##	enrollment_secondary_school		
## gdp	0.1564		
## gdp_growth	0.0179		
## gdp_per_capita	0.0000		
## u5mr	0.0000		
## foreign_direct_investment_net	0.4122		
## population_total	0.0690		

```
## population_growth          0.0089
## co2_emissions              0.1607
## cell_subs                  0.0000
## consumer_price_index       0.0003
## fertility_rate              0.3739
## enrollment_primary_secondary 0.0000
## enrollment_secondary_school
```

```
comb_cor = cor(t, method = c("spearman"))
corrplot(comb_cor, title="Coorelation High Income Group")
```



When you survey the high income country the correlations between specific variable is even lower. With the notable exception of population driving total gdp.

Conclusion

This initial analysis highlight some potential areas of exploration. It would be interesting to dig deeper into the relationships between gdp and other country attributes for specific groups of countries. It would also be interesting to dig deeper into how the region and income country groups impact the relationship between child mortality an other country attributes.