Final Project: Summary Statistics

David Aarhus

4/23/2020

This should include a summary table of means, max, mins, and standard deviations; data transformations performed for feature engineering, as well as at least five plots revealing interesting patterns to be studied. These must be motivated by the analysis – they cannot just be random plots.

#Libraries

library("tidyverse")

## ── Attaching packages ────────────────────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ ggplot2 3.2.1 ✓ purrr 0.3.3  
## ✓ tibble 2.1.3 ✓ dplyr 0.8.3  
## ✓ tidyr 1.0.0 ✓ stringr 1.4.0  
## ✓ readr 1.3.1 ✓ forcats 0.5.0

## ── Conflicts ───────────────────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library("ggplot2")  
library("caret")

## Loading required package: lattice

##   
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':  
##   
## lift

library("doBy")

rm(list = ls()) #removing all variables

#Loads in Data

suicide <- read.csv("/Users/DavidAarhus/Documents/310 R/Datasets/suicide.csv")  
names(suicide)

## [1] "country" "year" "sex"   
## [4] "age" "suicides\_no" "population"   
## [7] "suicides.100k.pop" "country.year" "HDI.for.year"   
## [10] "gdp\_for\_year...." "gdp\_per\_capita...." "generation"

#Data Transformation

#remove meaning less columns (HDI.for.year)  
suicide$HDI.for.year <- NULL  
suicide$country.year <- NULL  
#create new column (gdpM, gdpB, suicide\_hvy)  
suicide$gdpM <- suicide$gdp\_for\_year..../1e+6  
suicide$gdpB <- suicide$gdp\_for\_year..../1e+9  
suicide$suicide\_hvy <- ifelse(suicide$suicides\_no > 365, 1, 0)

#Summary tables

summary(suicide)

## country year sex age   
## Austria : 382 Min. :1985 female:13910 15-24 years:4642   
## Iceland : 382 1st Qu.:1995 male :13910 25-34 years:4642   
## Mauritius : 382 Median :2002 35-54 years:4642   
## Netherlands: 382 Mean :2001 5-14 years :4610   
## Argentina : 372 3rd Qu.:2008 55-74 years:4642   
## Belgium : 372 Max. :2016 75+ years :4642   
## (Other) :25548   
## suicides\_no population suicides.100k.pop gdp\_for\_year....   
## Min. : 0.0 Min. : 278 Min. : 0.00 Min. :4.692e+07   
## 1st Qu.: 3.0 1st Qu.: 97498 1st Qu.: 0.92 1st Qu.:8.985e+09   
## Median : 25.0 Median : 430150 Median : 5.99 Median :4.811e+10   
## Mean : 242.6 Mean : 1844794 Mean : 12.82 Mean :4.456e+11   
## 3rd Qu.: 131.0 3rd Qu.: 1486143 3rd Qu.: 16.62 3rd Qu.:2.602e+11   
## Max. :22338.0 Max. :43805214 Max. :224.97 Max. :1.812e+13   
##   
## gdp\_per\_capita.... generation gdpM   
## Min. : 251 Boomers :4990 Min. : 47   
## 1st Qu.: 3447 G.I. Generation:2744 1st Qu.: 8985   
## Median : 9372 Generation X :6408 Median : 48115   
## Mean : 16866 Generation Z :1470 Mean : 445581   
## 3rd Qu.: 24874 Millenials :5844 3rd Qu.: 260202   
## Max. :126352 Silent :6364 Max. :18120714   
##   
## gdpB suicide\_hvy   
## Min. : 0.047 Min. :0.0000   
## 1st Qu.: 8.985 1st Qu.:0.0000   
## Median : 48.115 Median :0.0000   
## Mean : 445.581 Mean :0.1281   
## 3rd Qu.: 260.202 3rd Qu.:0.0000   
## Max. :18120.714 Max. :1.0000   
##

summaryBy(suicides\_no ~ country, suicide, FUN = mean)

## country suicides\_no.mean  
## 1 Albania 7.462121e+00  
## 2 Antigua and Barbuda 3.395062e-02  
## 3 Argentina 2.210188e+02  
## 4 Armenia 6.392617e+00  
## 5 Aruba 6.011905e-01  
## 6 Australia 1.947528e+02  
## 7 Austria 1.310812e+02  
## 8 Azerbaijan 8.625000e+00  
## 9 Bahamas 3.369565e-01  
## 10 Bahrain 1.837302e+00  
## 11 Barbados 5.900000e-01  
## 12 Belarus 2.376667e+02  
## 13 Belgium 1.687124e+02  
## 14 Belize 1.035714e+00  
## 15 Bosnia and Herzegovina 1.325000e+01  
## 16 Brazil 6.091747e+02  
## 17 Bulgaria 1.010778e+02  
## 18 Cabo Verde 3.500000e+00  
## 19 Canada 3.090833e+02  
## 20 Chile 1.099328e+02  
## 21 Colombia 1.426882e+02  
## 22 Costa Rica 1.886667e+01  
## 23 Croatia 7.033969e+01  
## 24 Cuba 1.438125e+02  
## 25 Cyprus 2.314607e+00  
## 26 Czech Republic 1.356739e+02  
## 27 Denmark 5.794318e+01  
## 28 Dominica 0.000000e+00  
## 29 Ecuador 5.553763e+01  
## 30 El Salvador 4.056597e+01  
## 31 Estonia 2.791270e+01  
## 32 Fiji 2.303030e+00  
## 33 Finland 9.677299e+01  
## 34 France 9.142417e+02  
## 35 Georgia 1.221212e+01  
## 36 Germany 9.335321e+02  
## 37 Greece 3.324731e+01  
## 38 Grenada 1.225806e-01  
## 39 Guatemala 2.263611e+01  
## 40 Guyana 1.142000e+01  
## 41 Hungary 2.383581e+02  
## 42 Iceland 2.900524e+00  
## 43 Ireland 3.492778e+01  
## 44 Israel 3.036022e+01  
## 45 Italy 3.550000e+02  
## 46 Jamaica 9.019608e-01  
## 47 Japan 2.169091e+03  
## 48 Kazakhstan 3.254679e+02  
## 49 Kiribati 4.015152e-01  
## 50 Kuwait 3.220000e+00  
## 51 Kyrgyzstan 4.195513e+01  
## 52 Latvia 5.067460e+01  
## 53 Lithuania 1.070191e+02  
## 54 Luxembourg 5.263441e+00  
## 55 Macau 2.250000e+00  
## 56 Maldives 1.666667e-01  
## 57 Malta 1.572581e+00  
## 58 Mauritius 1.019372e+01  
## 59 Mexico 2.987608e+02  
## 60 Mongolia 4.230000e+01  
## 61 Montenegro 3.933333e+00  
## 62 Netherlands 1.330707e+02  
## 63 New Zealand 4.133046e+01  
## 64 Nicaragua 2.795833e+01  
## 65 Norway 4.720000e+01  
## 66 Oman 9.166667e-01  
## 67 Panama 1.161000e+01  
## 68 Paraguay 1.476235e+01  
## 69 Philippines 1.185000e+02  
## 70 Poland 4.829792e+02  
## 71 Portugal 7.426235e+01  
## 72 Puerto Rico 2.430914e+01  
## 73 Qatar 3.224719e+00  
## 74 Republic of Korea 7.035753e+02  
## 75 Romania 2.178952e+02  
## 76 Russian Federation 3.733772e+03  
## 77 Saint Kitts and Nevis 0.000000e+00  
## 78 Saint Lucia 6.845238e-01  
## 79 Saint Vincent and Grenadines 4.133333e-01  
## 80 San Marino 1.111111e-01  
## 81 Serbia 1.119398e+02  
## 82 Seychelles 4.537037e-01  
## 83 Singapore 2.712097e+01  
## 84 Slovakia 5.089773e+01  
## 85 Slovenia 4.212302e+01  
## 86 South Africa 3.050417e+01  
## 87 Spain 2.693602e+02  
## 88 Sri Lanka 4.215227e+02  
## 89 Suriname 6.446429e+00  
## 90 Sweden 1.055726e+02  
## 91 Switzerland 1.040357e+02  
## 92 Thailand 3.312665e+02  
## 93 Trinidad and Tobago 1.246605e+01  
## 94 Turkey 1.206071e+02  
## 95 Turkmenistan 2.478161e+01  
## 96 Ukraine 9.522321e+02  
## 97 United Arab Emirates 8.638889e+00  
## 98 United Kingdom 3.677554e+02  
## 99 United States 2.779605e+03  
## 100 Uruguay 3.910119e+01  
## 101 Uzbekistan 1.318295e+02

summaryBy(gdpB ~ country, suicide, FUN = mean)

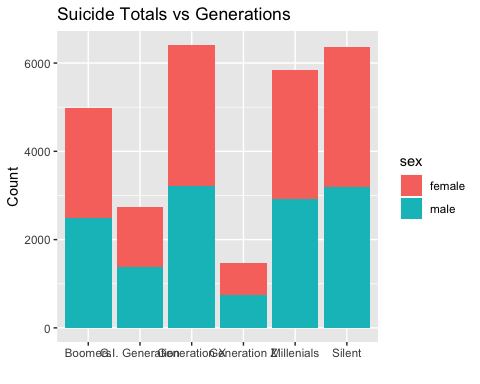
## country gdpB.mean  
## 1 Albania 5.211661e+00  
## 2 Antigua and Barbuda 8.035452e-01  
## 3 Argentina 2.742565e+02  
## 4 Armenia 5.386592e+00  
## 5 Aruba 2.196223e+00  
## 6 Australia 6.327501e+02  
## 7 Austria 2.660162e+02  
## 8 Azerbaijan 7.210687e+00  
## 9 Bahamas 7.613828e+00  
## 10 Bahrain 1.608647e+01  
## 11 Barbados 3.090574e+00  
## 12 Belarus 3.067641e+01  
## 13 Belgium 3.184024e+02  
## 14 Belize 9.560371e-01  
## 15 Bosnia and Herzegovina 1.860153e+01  
## 16 Brazil 1.022561e+03  
## 17 Bulgaria 2.714460e+01  
## 18 Cabo Verde 1.864824e+00  
## 19 Canada 9.131876e+02  
## 20 Chile 1.112543e+02  
## 21 Colombia 1.444638e+02  
## 22 Costa Rica 1.918353e+01  
## 23 Croatia 4.317814e+01  
## 24 Cuba 4.608462e+01  
## 25 Cyprus 2.137970e+01  
## 26 Czech Republic 1.220562e+02  
## 27 Denmark 2.530342e+02  
## 28 Dominica 9.858519e-02  
## 29 Ecuador 3.960132e+01  
## 30 El Salvador 1.354975e+01  
## 31 Estonia 1.444586e+01  
## 32 Fiji 2.927373e+00  
## 33 Finland 1.753008e+02  
## 34 France 1.781194e+03  
## 35 Georgia 7.776764e+00  
## 36 Germany 2.742233e+03  
## 37 Greece 1.766047e+02  
## 38 Grenada 5.762393e-01  
## 39 Guatemala 2.523613e+01  
## 40 Guyana 1.120870e+00  
## 41 Hungary 8.952380e+01  
## 42 Iceland 1.073664e+01  
## 43 Ireland 1.315586e+02  
## 44 Israel 1.394319e+02  
## 45 Italy 1.481752e+03  
## 46 Jamaica 7.954517e+00  
## 47 Japan 4.339221e+03  
## 48 Kazakhstan 7.868722e+01  
## 49 Kiribati 5.928353e-02  
## 50 Kuwait 7.217026e+01  
## 51 Kyrgyzstan 3.309952e+00  
## 52 Latvia 1.831660e+01  
## 53 Lithuania 2.771062e+01  
## 54 Luxembourg 3.059349e+01  
## 55 Macau 6.265844e+00  
## 56 Maldives 1.533642e+00  
## 57 Malta 5.291359e+00  
## 58 Mauritius 6.028693e+00  
## 59 Mexico 6.803078e+02  
## 60 Mongolia 1.118346e+01  
## 61 Montenegro 2.457459e+00  
## 62 Netherlands 5.425437e+02  
## 63 New Zealand 8.269044e+01  
## 64 Nicaragua 1.075657e+01  
## 65 Norway 2.525657e+02  
## 66 Oman 6.270247e+01  
## 67 Panama 2.020910e+01  
## 68 Paraguay 1.168725e+01  
## 69 Philippines 1.136853e+02  
## 70 Poland 2.944583e+02  
## 71 Portugal 1.389200e+02  
## 72 Puerto Rico 6.322157e+01  
## 73 Qatar 1.100206e+02  
## 74 Republic of Korea 6.734208e+02  
## 75 Romania 9.608785e+01  
## 76 Russian Federation 8.843229e+02  
## 77 Saint Kitts and Nevis 2.160186e-01  
## 78 Saint Lucia 8.409455e-01  
## 79 Saint Vincent and Grenadines 4.884483e-01  
## 80 San Marino 1.425417e+00  
## 81 Serbia 3.112522e+01  
## 82 Seychelles 8.564692e-01  
## 83 Singapore 1.254921e+02  
## 84 Slovakia 5.379573e+01  
## 85 Slovenia 3.590006e+01  
## 86 South Africa 2.494661e+02  
## 87 Spain 8.565680e+02  
## 88 Sri Lanka 1.532645e+01  
## 89 Suriname 1.943388e+00  
## 90 Sweden 3.549738e+02  
## 91 Switzerland 4.565306e+02  
## 92 Thailand 2.095240e+02  
## 93 Trinidad and Tobago 1.044825e+01  
## 94 Turkey 8.382300e+02  
## 95 Turkmenistan 1.157956e+01  
## 96 Ukraine 8.389111e+01  
## 97 United Arab Emirates 2.532432e+02  
## 98 United Kingdom 1.816067e+03  
## 99 United States 1.051071e+04  
## 100 Uruguay 2.337827e+01  
## 101 Uzbekistan 2.286003e+01

summaryBy(suicides\_no ~ sex, suicide, FUN = mean)

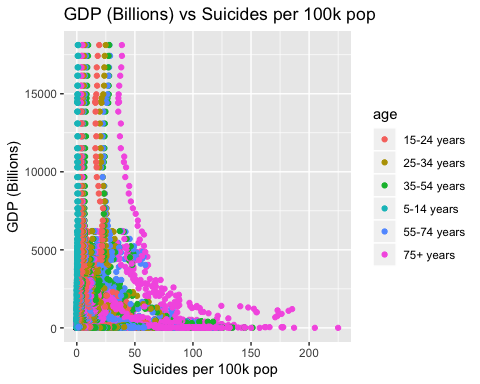
## sex suicides\_no.mean  
## 1 female 112.1143  
## 2 male 373.0345

#GGPLOTS

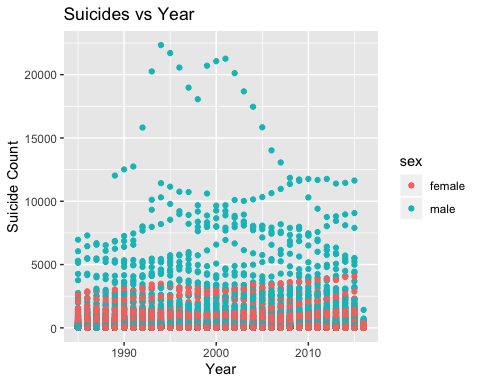
ggplot(suicide, aes(generation)) +  
 geom\_bar(aes(fill = sex)) +  
 labs(x = " ", y = "Count", title = "Suicide Totals vs Generations")



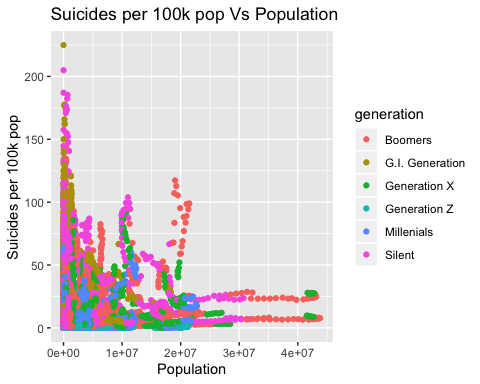
ggplot(suicide, aes(suicides.100k.pop, gdpB)) +  
 geom\_point(aes(color = age)) +  
 labs(x = "Suicides per 100k pop ", y = "GDP (Billions)", title = "GDP (Billions) vs Suicides per 100k pop")



ggplot(suicide, aes(year, suicides\_no)) +  
 geom\_point(aes(color = sex)) +  
 labs(x = "Year", y = "Suicide Count", title = "Suicides vs Year")



ggplot(suicide, aes(population, suicides.100k.pop )) +  
 geom\_point(aes(color = generation)) +  
 labs(x = "Population", y = "Suicides per 100k pop", title = "Suicides per 100k pop Vs Population")



ggplot(suicide, aes(year, suicides.100k.pop)) +  
 geom\_point(aes(color = generation)) +  
 labs(x = "Year", y = "Suicides per 100k pop", title = "Suicides per 100k pop Vs Year")

