Viz\_OneA

D21124026

2022-12-09

## R Markdown

# -------------------------------------------------- #  
# -------------------------------------------------- #  
  
# Data Visualisation Assignment 2  
# Visualisations in R  
#  
# Student No. d21124026  
# Name : Ciaran Finnegan  
# TU060 Data Science MSc  
#  
# December 2022  
  
  
# Visualisation One - Comparison of Unemployment Rates   
# Per County in Ireland in Census 2011 data   
# -------------------------------------------------- #  
# -------------------------------------------------- #  
  
  
# load required libraries  
library(rgeos)

library(maptools)

library(curl)

library(readr)

library(ggplot2)

library(sqldf)

library(tidyverse)

library(viridis)

library(dplyr)

# There were challenges with accessing data file locations for all datasets.  
# To allow the R code to run anywhere the required data files have been hosted   
# on GitHub where required.  
  
# Build URL connecting string to download OSI country boundaries data file   
# for counties in the Republic of Ireland  
sGitHub\_Datasource1 <-"https://github.com/JackDaedalus/DataVizLabs/raw/"  
sGitHub\_Datasource2 <- paste(sGitHub\_Datasource1,"dfa3d486a5ea74a588e9768141b35f570eff3c57/CA2/", sep = "", collapse=NULL)  
sGitHub\_Datafile <- "Counties\_-\_OSi\_National\_Statutory\_Boundaries\_-\_2019\_-\_Generalised\_20m.zip"  
sGitHub\_Datasource <- paste(sGitHub\_Datasource2,sGitHub\_Datafile, sep = "", collapse=NULL)  
  
county\_map\_source <- sGitHub\_Datasource  
  
# Download zip file from from GitHub and extract 2019 OSI County data for 26 Irish counties  
temp\_1 <- tempfile()  
temp\_2 <- tempfile()  
source <- county\_map\_source  
temp\_1 <- curl\_download(url = source, destfile = temp\_1, quiet = FALSE)  
unzip(temp\_1, exdir = temp\_2)  
  
  
  
  
# Read the shape file extracted from the downloaded zip file  
spdf <- readShapePoly(file.path(temp\_2,"Counties\_\_\_OSi\_National\_Statutory\_Boundaries\_\_\_Generalised\_20m.shp"))

# Generate a new dataframe that will work with ggplot to generate Ireland county map  
spdf@data$id <- rownames(spdf@data)  
spdf.points <- fortify(spdf, region="id")  
counties <- inner\_join(spdf.points, spdf@data, by="id")

# ----------------------------------------------------------#  
# Load Census Theme Data for 2011 for Irish counties  
# ----------------------------------------------------------#  
  
  
# This data is available directly from the CSO website  
# Select only the required unemployment data   
# Rename the columns to increase understanding of the data  
df2011CountyThemes <- read\_delim("https://www.cso.ie/en/media/csoie/census/documents/saps2011files/AllThemesTablesCTY.csv",show\_col\_types = FALSE) %>%  
 select(GEOGID, GEOGDESC, T8\_1\_LFFJT, T8\_1\_ULGUPJT, T8\_1\_TT) %>%  
 rename(Looking\_for\_Work = T8\_1\_LFFJT, Unemployed = T8\_1\_ULGUPJT, Total\_Workforce = T8\_1\_TT)  
  
  
  
# ----------------------------------------------------------------#  
# The Census theme data breaks down the counties in certain cases   
# for metropolitan areas. This data needs to be re-merged to match   
# the county boundaries in the OSI dataframe  
# ----------------------------------------------------------------#  
  
  
# Start with Dublin...  
df2011DublinThemes <- df2011CountyThemes %>%  
 filter(GEOGID %in% c("C02","C03","C04","C05"))  
  
# Group the country regions and sum all unemployment data for Dublin overall  
df2011DublinThemesTotal <- sqldf("Select 'C35' as GEOGID,  
 'Dublin' as Dublin,  
 sum(Looking\_for\_Work),  
 sum(Unemployed),  
 sum(Total\_Workforce)  
 from df2011DublinThemes  
 group by Dublin")  
  
  
  
  
# The Cork areas are combined next...  
df2011CorkThemes <- df2011CountyThemes %>%  
 filter(GEOGID %in% c("C17","C18"))  
  
# Group the country regions and sum all unemployment data for Dublin overall  
df2011CorkThemesTotal <- sqldf("Select 'C36' as GEOGID,  
 'Cork' as Cork,  
 sum(Looking\_for\_Work),  
 sum(Unemployed),  
 sum(Total\_Workforce)  
 from df2011CorkThemes  
 group by Cork")  
  
  
  
  
# The Limerick areas are combined next...  
df2011LimerickThemes <- df2011CountyThemes %>%  
 filter(GEOGID %in% c("C20","C21"))  
  
# Group the country regions and sum all unemployment data for Limerick overall  
df2011LimerickThemesTotal <- sqldf("Select 'C36' as GEOGID,  
 'Limerick' as Limerick,  
 sum(Looking\_for\_Work),  
 sum(Unemployed),  
 sum(Total\_Workforce)  
 from df2011LimerickThemes  
 group by Limerick")  
  
  
  
  
# The Tipperary areas are combined next...  
df2011TipperaryThemes <- df2011CountyThemes %>%  
 filter(GEOGID %in% c("C22","C23"))  
  
# Group the country regions and sum all unemployment data for Tipperary overall  
df2011TipperaryThemesTotal <- sqldf("Select 'C37' as GEOGID,  
 'Tipperary' as Tipperary,  
 sum(Looking\_for\_Work),  
 sum(Unemployed),  
 sum(Total\_Workforce)  
 from df2011TipperaryThemes  
 group by Tipperary")  
  
  
  
  
# The Waterford areas are combined next...  
df2011WaterfordThemes <- df2011CountyThemes %>%  
 filter(GEOGID %in% c("C24","C25"))  
  
# Group the country regions and sum all unemployment data for Waterford overall  
df2011WaterfordThemesTotal <- sqldf("Select 'C38' as GEOGID,  
 'Waterford' as Waterford,  
 sum(Looking\_for\_Work),  
 sum(Unemployed),  
 sum(Total\_Workforce)  
 from df2011WaterfordThemes  
 group by Waterford")  
  
  
  
  
# The Galway areas are combined next...  
df2011GalwayThemes <- df2011CountyThemes %>%  
 filter(GEOGID %in% c("C26","C27"))  
  
# Group the country regions and sum all unemployment data for Galway overall  
df2011GalwayThemesTotal <- sqldf("Select 'C39' as GEOGID,  
 'Galway' as Galway,  
 sum(Looking\_for\_Work),  
 sum(Unemployed),  
 sum(Total\_Workforce)  
 from df2011GalwayThemes  
 group by Galway")  
  
  
  
  
  
  
# Add Collated County data to revised county theme data  
df2011CountyThemes <- data.frame(rbind(as.matrix(df2011CountyThemes), as.matrix(df2011DublinThemesTotal))) # Dublin  
df2011CountyThemes <- data.frame(rbind(as.matrix(df2011CountyThemes), as.matrix(df2011CorkThemesTotal))) # Cork  
df2011CountyThemes <- data.frame(rbind(as.matrix(df2011CountyThemes), as.matrix(df2011LimerickThemesTotal))) # Limerick  
df2011CountyThemes <- data.frame(rbind(as.matrix(df2011CountyThemes), as.matrix(df2011TipperaryThemesTotal))) # Tipperary  
df2011CountyThemes <- data.frame(rbind(as.matrix(df2011CountyThemes), as.matrix(df2011WaterfordThemesTotal))) # Waterford  
df2011CountyThemes <- data.frame(rbind(as.matrix(df2011CountyThemes), as.matrix(df2011GalwayThemesTotal))) # Galway

# ----------------------------------------------------------------#  
# Additional data manipulation routines to reformat data and add  
# calculated fields to 2011 dataframe  
# ----------------------------------------------------------------#  
  
  
# Convert county names to upper case to match map dataframe  
df2011CountyThemes <- mutate\_all(df2011CountyThemes, .funs=toupper)  
  
  
# Reconvert County unemployment data columns back to numeric  
df2011CountyThemes$Looking\_for\_Work = as.numeric(as.character(df2011CountyThemes$Looking\_for\_Work))  
df2011CountyThemes$Unemployed = as.numeric(as.character(df2011CountyThemes$Unemployed))  
df2011CountyThemes$Total\_Workforce = as.numeric(as.character(df2011CountyThemes$Total\_Workforce))  
  
  
  
# Remove the redundant county sub-breakdowns for unemployment data  
df2011CountyThemes <- df2011CountyThemes %>%  
 filter(!GEOGID %in% c("C02","C03","C04","C05","C17","C18","C20","C21","C22","C23","C24","C25","C26","C27")) %>%  
 rename(COUNTY = GEOGDESC)  
  
  
  
# Calculate Unemployment rate by County and add to Dataframe  
df2011CountyThemes$Unemploy\_Rate <- ((df2011CountyThemes$Looking\_for\_Work + df2011CountyThemes$Unemployed) / df2011CountyThemes$Total\_Workforce) \* 100  
  
  
  
  
# Using cut() function to create 2011 categorical bands for rates of unemployment  
df2011CountyThemes$Unemploy\_Pct <- cut(df2011CountyThemes$Unemploy\_Rate,   
 breaks = c(0, 6.99, 9.99, 10.99, 11.99, 12.99, 13.99, 14.99, 99),   
 labels = c("<7%", "7-10%", "10-11%", "11-12%", "12-13%", "13-14%", "14-15%", "15%+"))  
  
  
  
# Join dataframe on county names  
dfCountyMap <- left\_join(counties, df2011CountyThemes, by = "COUNTY")  
  
# Set up factor in dataframe for visualisation  
dfCountyMap$COUNTY <- factor(dfCountyMap$COUNTY)

#---------------------------------------------------------------------------------#  
# Generate the 2011 Country Map Plot with Unemployment Rates  
#---------------------------------------------------------------------------------#  
  
ggplot(dfCountyMap) +   
 geom\_polygon(colour="black", aes(x=long, y=lat, group=group, fill=Unemploy\_Pct)) +   
 labs(x = NULL, y = NULL,   
 title = "Unemployment Rate by County (2011) for Ireland",  
 subtitle = "Sources: Census 2011",   
 caption = "Plot by C.Finegan d21124026") +   
 theme(axis.line=element\_blank(),   
 axis.ticks=element\_blank(),   
 axis.text=element\_blank(),  
 axis.title=element\_blank(),  
 panel.grid = element\_blank(),  
 plot.caption.position = 'plot',  
 plot.title.position = 'plot',  
 legend.position = "bottom") +   
 labs(fill = "Unemployment Rate (%)") +  
 # Colour scale for unemployment rate  
 scale\_fill\_brewer(palette="YlOrRd")

