Developing Data Products - Week 3 Assignment

*Surbhi Agarwal*

*21 JULY 2020*

knitr::opts\_chunk$set(echo = TRUE, cache = TRUE)

Introduction

* This project was created as part of the Developing Data Products course of the Coursera [Data Science Specialisation](https://www.coursera.org/specializations/jhu-data-science).
* The goal of the project is to create a web page presentation using R Markdown that features a plot created with Plotly, and to host the resulting web page on either GitHub Pages, RPubs, or NeoCities.
* The interactive plot on the next slide represents the number of road accidents in Great Britain from 2005 to 2015, grouped by severity (slight, serious, or fatal).

rm(list=ls())

**library**(plotly)

**library**(data.table)

**library**(tidyr)

**library**(lubridate)

**library**(zoo)

*# The source data sets are not included in this repository.*

*# To reproduce this presentation, you will first need to download the two*

*# following zipped data sets:*

*# - All STATS19 data (accident, casualties and vehicle tables) for 2005 to*

*# 2014", from*

*# https://data.gov.uk/dataset/road-accidents-safety-data/resource/8ecee6ac-33fd-4f5b-8973-e900cc65d24a)*

*# - Road Safety - Accidents 2015, from*

*# https://data.gov.uk/dataset/road-accidents-safety-data/resource/ceb00cff-443d-4d43-b17a-ee13437e9564)*

*# Then extract the `Accidents0514.csv` and `Accidents\_2015.csv` files from*

*# the zip files in a subdirectory named `data`.*

*# read data for 2005-2014 and 2015 as data tables and keep only severity and*

*# date columns*

accidents0514 <- fread("data/Accidents0514.csv", header = TRUE, sep = ",")

##

Read 0.0% of 1640597 rows

Read 9.1% of 1640597 rows

Read 17.7% of 1640597 rows

Read 26.2% of 1640597 rows

Read 28.6% of 1640597 rows

Read 37.8% of 1640597 rows

Read 43.3% of 1640597 rows

Read 53.0% of 1640597 rows

Read 62.8% of 1640597 rows

Read 65.8% of 1640597 rows

Read 78.0% of 1640597 rows

Read 87.2% of 1640597 rows

Read 97.5% of 1640597 rows

Read 1640597 rows and 32 (of 32) columns from 0.210 GB file in 00:00:17

accidents0514 <- accidents0514 %>%

select(Accident\_Severity, Date)

accidents15 <- fread("data/Accidents\_2015.csv", header = TRUE, sep = ",")

accidents15 <- accidents15 %>%

select(Accident\_Severity, Date)

*# concatenate data tables and free up environment*

accidents <- rbind(accidents0514, accidents15)

rm(list = c("accidents0514", "accidents15"))

*# convert severity to factor and add labels*

accidents$Accident\_Severity <- factor(accidents$Accident\_Severity, levels = 1:3, labels = c("Fatal", "Serious", "Slight"))

*# convert date strings to Date objects*

accidents$Date <- dmy(accidents$Date)

*# group data by date and severity, get count, one row per date*

accident\_count <- accidents %>%

group\_by(Date, Accident\_Severity) %>%

summarise(count = n()) %>%

spread(key = Accident\_Severity, value = count) %>%

as.data.frame()

*# create a smoother for each severity to visualise general trends*

loess\_slight <- loess(Slight ~ as.numeric(Date), data = accident\_count)

loess\_serious <- loess(Serious ~ as.numeric(Date), data = accident\_count)

loess\_fatal <- loess(Fatal ~ as.numeric(Date), data = accident\_count)

Road accidents in Great Britain (2005-2015)

*# plot data*

plot\_ly(accident\_count) %>%

add\_trace(x = ~Date, y = ~Slight, type="scatter", mode = "markers",

name = "slight", legendgroup = "slight",

marker = list(color = "#52A9BD")) %>%

add\_trace(x = ~Date, y = ~Serious, type="scatter", mode = "markers",

name = "serious", legendgroup = "serious",

marker = list(color = "#FFF16B")) %>%

add\_trace(x = ~Date, y = ~Fatal, type="scatter", mode = "markers",

name = "fatal", legendgroup = "fatal",

marker = list(color = "#F5677D")) %>%

add\_trace(x = as.Date(loess\_slight$x), y = fitted(loess\_slight),

type="scatter", mode = "lines",

line = list(color = '#1A7A90'),

name = "slight Loess smoother", legendgroup = "slight",

hoverinfo = 'none', showlegend = FALSE) %>%

add\_trace(x = as.Date(loess\_serious$x), y = fitted(loess\_serious),

type="scatter", mode = "lines",

line = list(color = '#E9D625'),

name = "serious Loess smoother", legendgroup = "serious",

hoverinfo = 'none', showlegend = FALSE) %>%

add\_lines(x = as.Date(loess\_fatal$x), y = fitted(loess\_fatal),

type="scatter", mode = "lines",

line = list(color = '#DC2340'),

name = "fatal Loess smoother", legendgroup = "fatal",

hoverinfo = 'none', showlegend = FALSE) %>%

layout(xaxis = list(title = "Year"),

yaxis = list(title = "Number of Accidents")

)

## Warning: Ignoring 39 observations

2006200820102012201420160100200300400500600700

slightseriousfatalYearNumber of Accidents