MATH2270 Assignment 2

Visualising Open Data

Student Details

- Student Name (s3000000)

```
## Load Libraries
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
require(dplyr)
```

```
## Loading required package: dplyr
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
require(splitstackshape)
```

```
## Loading required package: splitstackshape
```

```
## Loading required package: data.table
```

```
## -----------------------------------------------------------------------
```

```
## data.table + dplyr code now lives in dtplyr.
## Please library(dtplyr)!
```

```
## -----------------------------------------------------------------------
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
library(googleVis)
```

```
## Creating a generic function for 'toJSON' from package 'jsonlite' in package 'googleVis'
```

```
##
## Welcome to googleVis version 0.6.2
##
## Please read Google's Terms of Use
## before you start using the package:
## https://developers.google.com/terms/
##
## Note, the plot method of googleVis will by default use
```

```
## the standard browser to display its output.
##
## See the googleVis package vignettes for more details,
## or visit http://github.com/mages/googleVis.
##
## To suppress this message use:
## suppressPackageStartupMessages(library(googleVis))
```

```
require(ggvis)
```

```
## Loading required package: ggvis
```

```
##
## Attaching package: 'ggvis'
```

```
## The following object is masked from 'package:ggplot2':
##
##     resolution
```

Data

```
# Load your data and prepare for visualisation
Data2017 = read.csv('data/survey_results_public.csv')

Education = Data2017 %>% select(Respondent,EducationTypes,SelfTaughtTypes,DeveloperType) %>%
  filter(!is.na(EducationTypes) & !is.na(EducationTypes) & DeveloperType %in%
           c('Data scientist','Machine learning specialist','Developer with a statistics or mathematics

Education_long <- cSplit(indt = Education,
                         splitCols = c('EducationTypes','SelfTaughtTypes','DeveloperType'), sep = ";", d

x = data.frame(table(Education_long$DeveloperType,Education_long$EducationTypes))
y = data.frame(table(Education_long$EducationTypes,Education_long$SelfTaughtTypes))

defaulter<-rbind(x,y)
```

Visualisation

```
# Plotting

sk2 <- gvisSankey(defaulter, from='Var1', to='Var2', weight='Freq',
                  options=list(height=400, width=600))
plot(sk2)
```

```
## starting httpd help server ... done
```

Data Source https://insights.stackoverflow.com/survey/