Final_Project

Kyle Calcagno, Daehyun Suk, Matt Watson

5/4/2018

```
library(readr)
beers = read_csv("beers.csv")
```

```
## Warning in rbind(names(probs), probs_f): number of columns of result is not
## a multiple of vector length (arg 1)
```

```
## Warning: 63 parsing failures.
## row # A tibble: 5 x 5 col     row col        expected               actual                     file
## ... ................. ... ...........................................................................
## See problems(...) for more details.
```

```
categories = read_csv("categories.csv")


styles = read_csv("styles.csv")


breweries = read_csv("breweries.csv")


brewerie_geocode = read_csv("breweries_geocode.csv")
```

```
beers = beers[ , -c(7:12)]
```

```
beers = beers[ ! beers$id %in% NA, ]
```

```
beers = beers[ ! beers$cat_id %in% c(0, -1), ]  #Remove cat_id of 0 and -1
#dim(beers)               #4398 Variables Remain
```

```
beers = beers[ ! beers$brewery_id %in% c(1416, 1417, 1419, 1420, 1421, 1422, 1423), ]
```

```
colnames(styles)[colnames(styles) == "id"] = "style_id"

styles_name = subset(styles, select = c(style_id, style_name))

beers = merge(beers, styles_name, by = "style_id", all =TRUE)
```

```
colnames(breweries)[colnames(breweries) == "id"] = "brewery_id"

breweries_countries = subset(breweries, select = c(brewery_id, country))

beers = merge(beers, breweries_countries, by = "brewery_id", all = TRUE)
```

```
colnames(categories)[colnames(categories) == "id"] = "cat_id"

categories_names = subset(categories, select = c(cat_id, cat_name))

beers = merge(beers, categories_names, by = "cat_id", all = TRUE)
```

```
latlong = subset(brewerie_geocode, select = c(brewery_id, latitude, longitude))
beers = merge(beers, latlong, by = "brewery_id", all = TRUE)
```

```
beers = beers[ ! beers$id %in% NA, ]
beers = beers[ ! beers$latitude %in% NA, ]
```

```
write.csv(beers, "beers_cleaned.csv")
```