

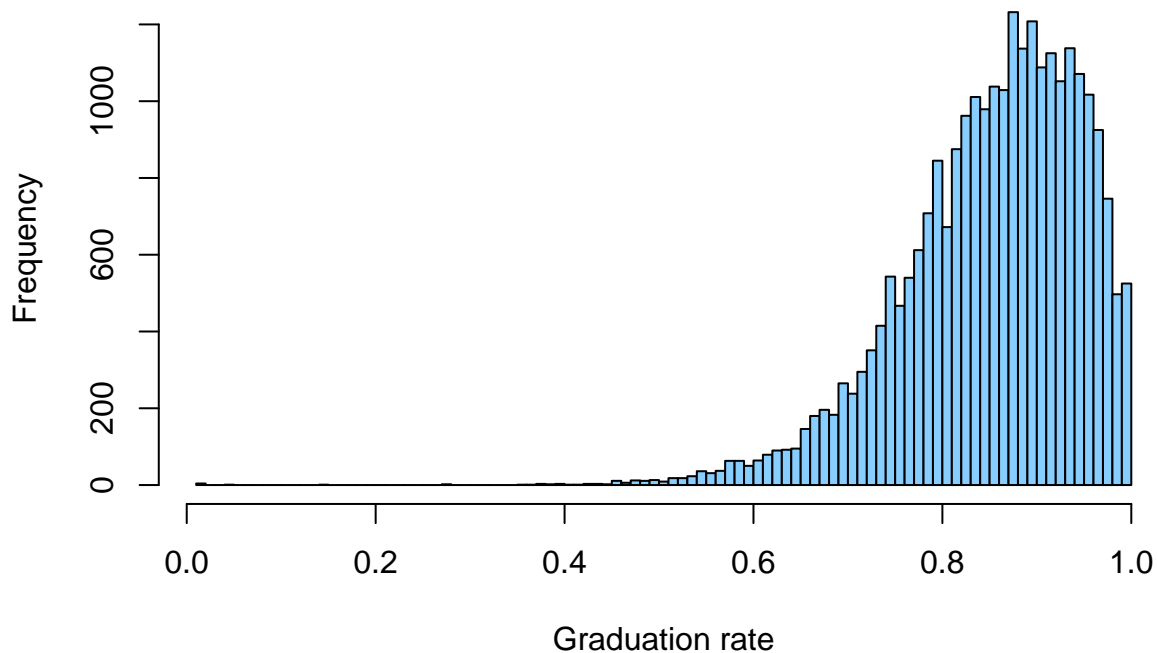
# NY grad rate analysis

Nick Orangio

10/27/2020

```
###  
# import data and look at grad rates  
###  
  
# CORNELL DATA  
# import cornell data  
corn_df <- read.csv("/Users/nickorangio/NYC_hs_grad/CSE6242/Data/apm_demo_faru_grad_clean.csv")  
# str(corn_df)  
  
# engineer grad rate column  
corn_df <- transform(corn_df, grad_rate = gr_graduated / total)  
  
# drop zero and negative grad rates  
corn_df <- corn_df[corn_df$grad_rate > 0, ]  
  
# view distribution of grad rates from cornell data  
hist(corn_df$grad_rate, col = "skyblue1", breaks = 100, main = "Histogram of graduation rates by cohort",  
      xlab = "Graduation rate")
```

**Histogram of graduation rates by cohort for all students (Cornell)**



```

# NYSED Data
# import combined grad rate tables from NYSED
nysed_df <- read.csv("/Users/nickorangio/NYC_hs_grad/combined_grad.csv")

# filter for just subgroup = 1
nysed_df1 <- nysed_df[nysed_df$subgroup_code == 1, ]

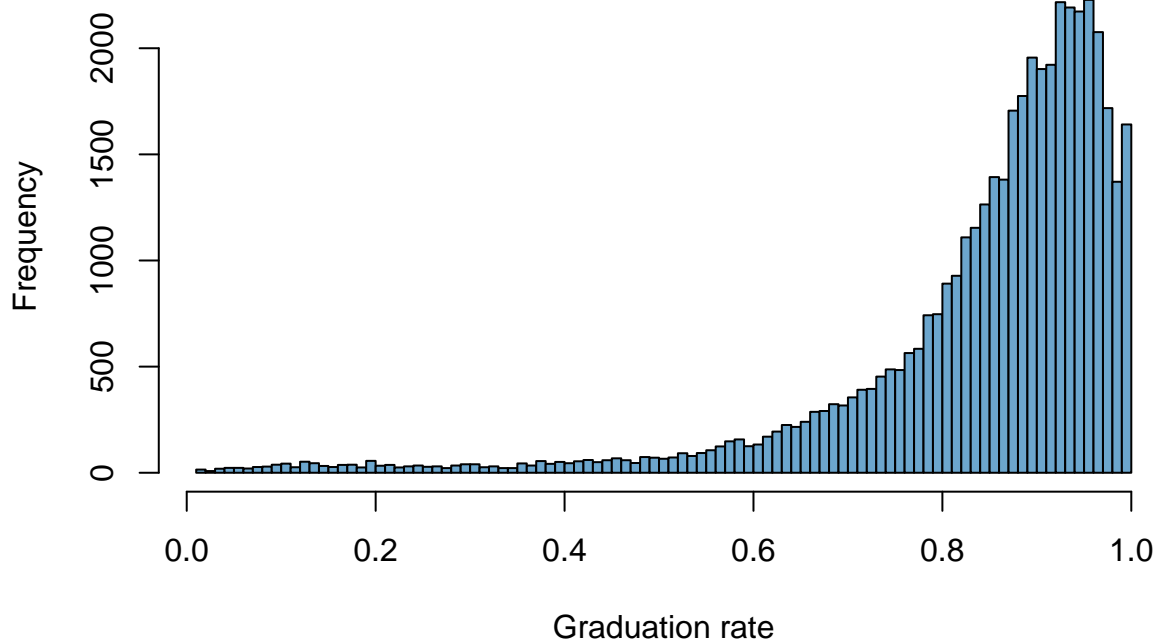
# filter out one percent and under grad rates
nysed_df1 <- nysed_df1[nysed_df1$grad_pct > 1, ]

# convert grad percentage to numeric
nysed_df1$grad_pct <- as.numeric(sub("%", "", nysed_df1$grad_pct)) / 100

# histogram of grad percent
hist(nysed_df1$grad_pct, col = "skyblue3", breaks = 100, main = "Histogram of graduation rates by cohort",
      xlab = "Graduation rate")

```

## Histogram of graduation rates by cohort for all students (NYSED)



```

###
# correlation analysis for Cornell data
###

library(corrplot)

## Warning: package 'corrplot' was built under R version 4.0.2
## corrplot 0.84 loaded

# extract columns for correlations
cor_data <- corn_df[ , c("rev_pupil", "exp_pupil", "lrev", "tsal", "pps", "cds", "boc", "staid", "grad_

# run correlations

```

```
cor_ob <- cor(cor_data)
corrplot.mixed(cor_ob, tl.cex = 0.7, order = "hclust")
```

