# Modeling 2 Final

Alex McGraw, Justin Wagers, Christian Lee, Joe Niehaus

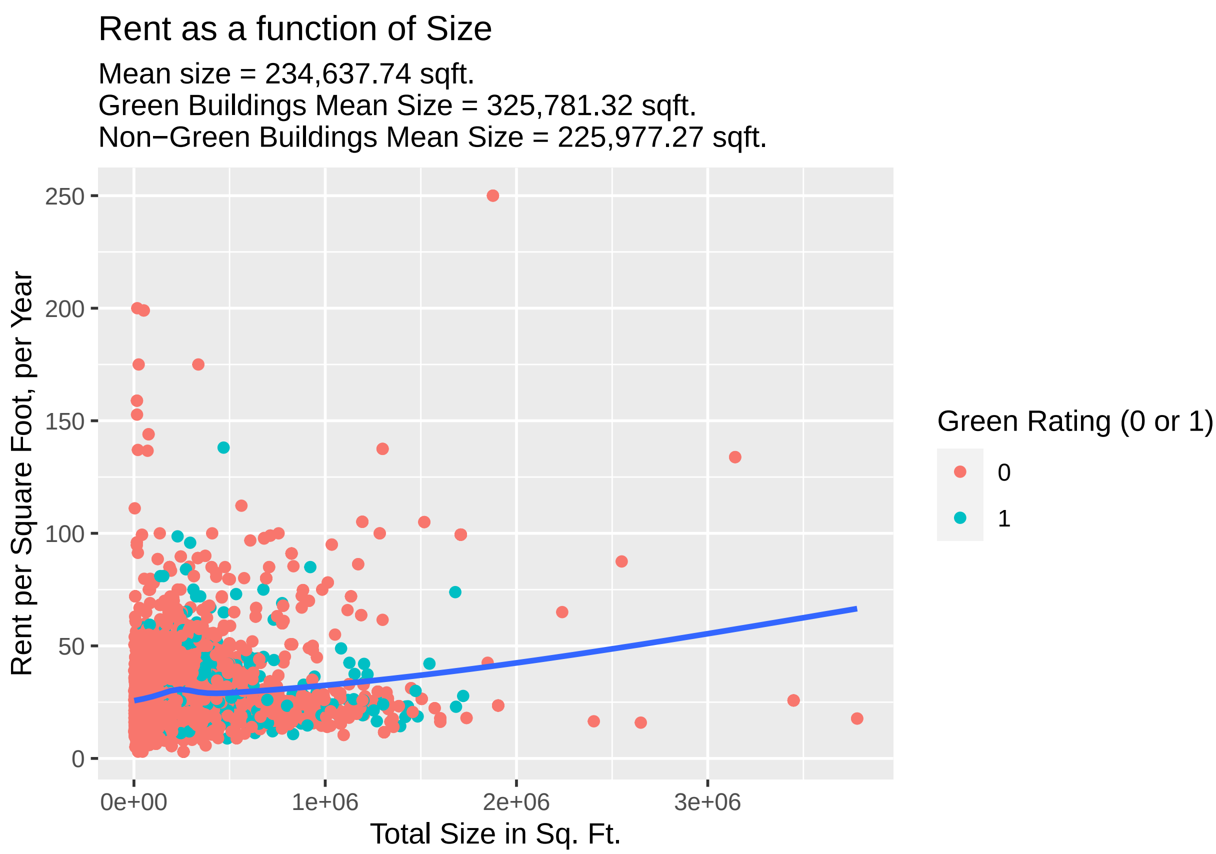
8/18/2020

## Question 1 - Green Buildings

### 1

From the beginning, I'm suspicious of the excel guru's analysis because first, he cut out buildings that had less than 10% occupancy. This did not take into account the age of the building. If it was developed in the last six months or a year, this less than 10% rule should not be applicable to the building. Yes, there are cases where this less than 10% metric could be a telling factor of the quality of the building or administration, it should not be used as a blanket statement that <=10% occupancy is bad. Secondly, he just took an average of the rent of non-green buildings and green buildings to then forecast premium rent rates of green buildings. He completely disregarded the 21 other variables that could have been used.

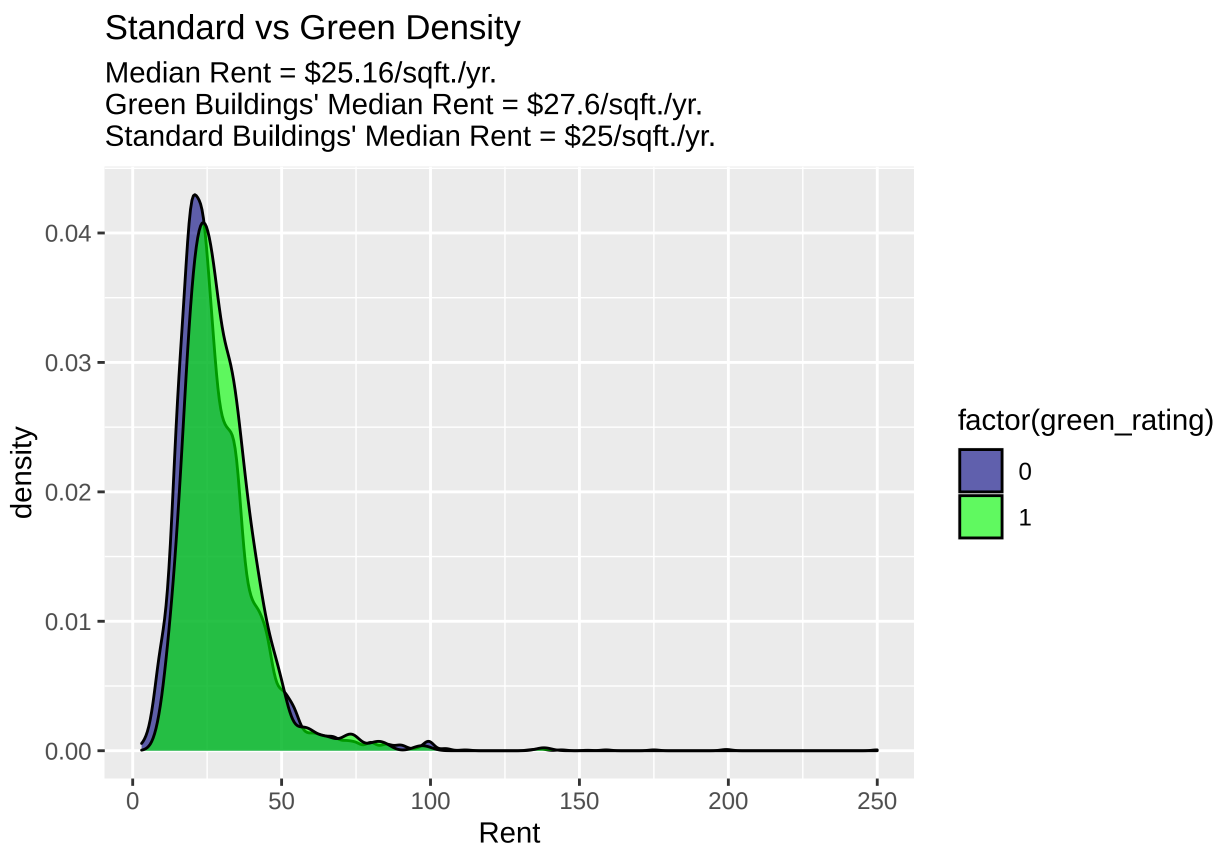
```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




This initial plot shows how Rent changes as the total size of the building increases.While the fitted line is not a good predictor, we can see that the rent price does increase as size of the building increases.

Green buildings have a higher mean building size than standard ones, which can be a reason to believe that green buildings are able to fetch higher rent prices.

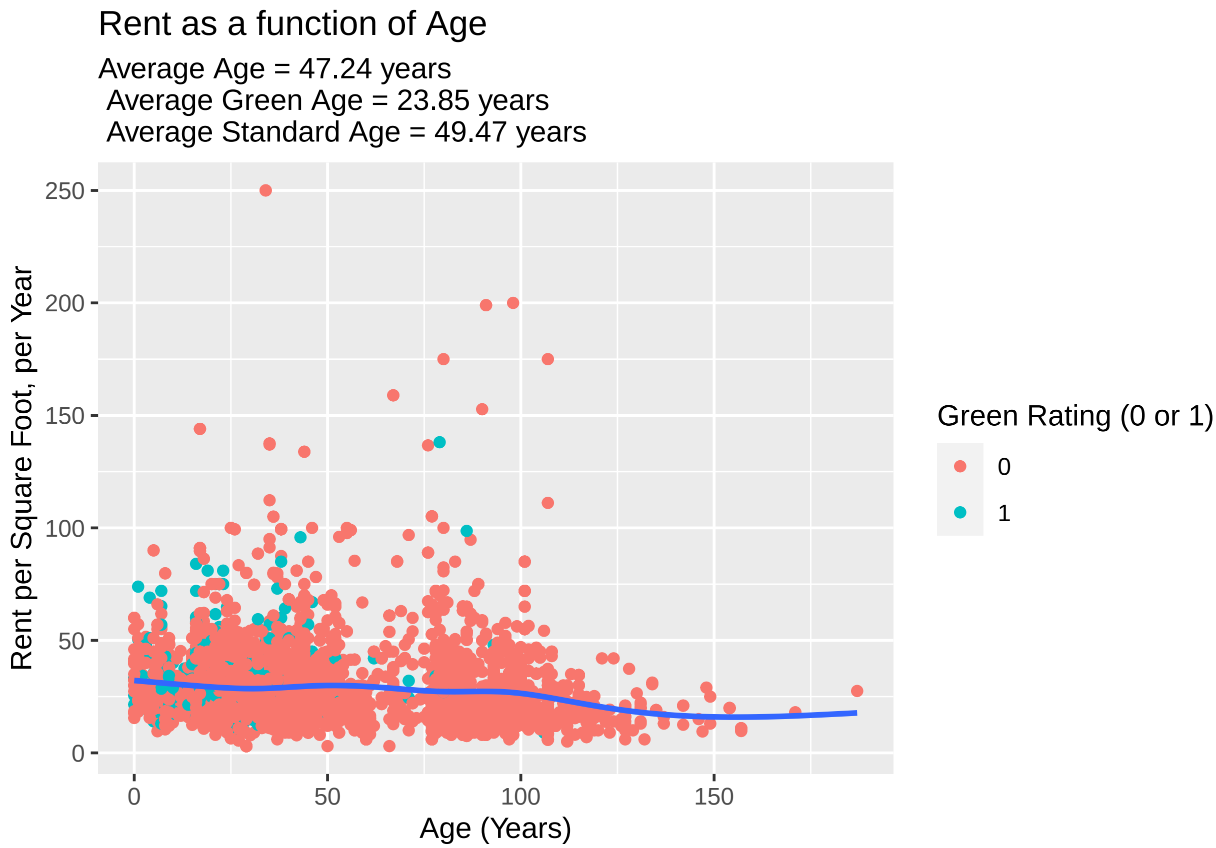
**2**




Here we can see that while green buildings have a slightly higher median compared to standard buildings, they are practically identical. This slightly higher median rent could be attributed to the on average larger building size of the green buildings, but we must look into more variables to get a better idea.

**3**

```
## ‘geom_smooth()‘ using method = ’gam’ and formula ’y ~ s(x, bs = "cs")’
```

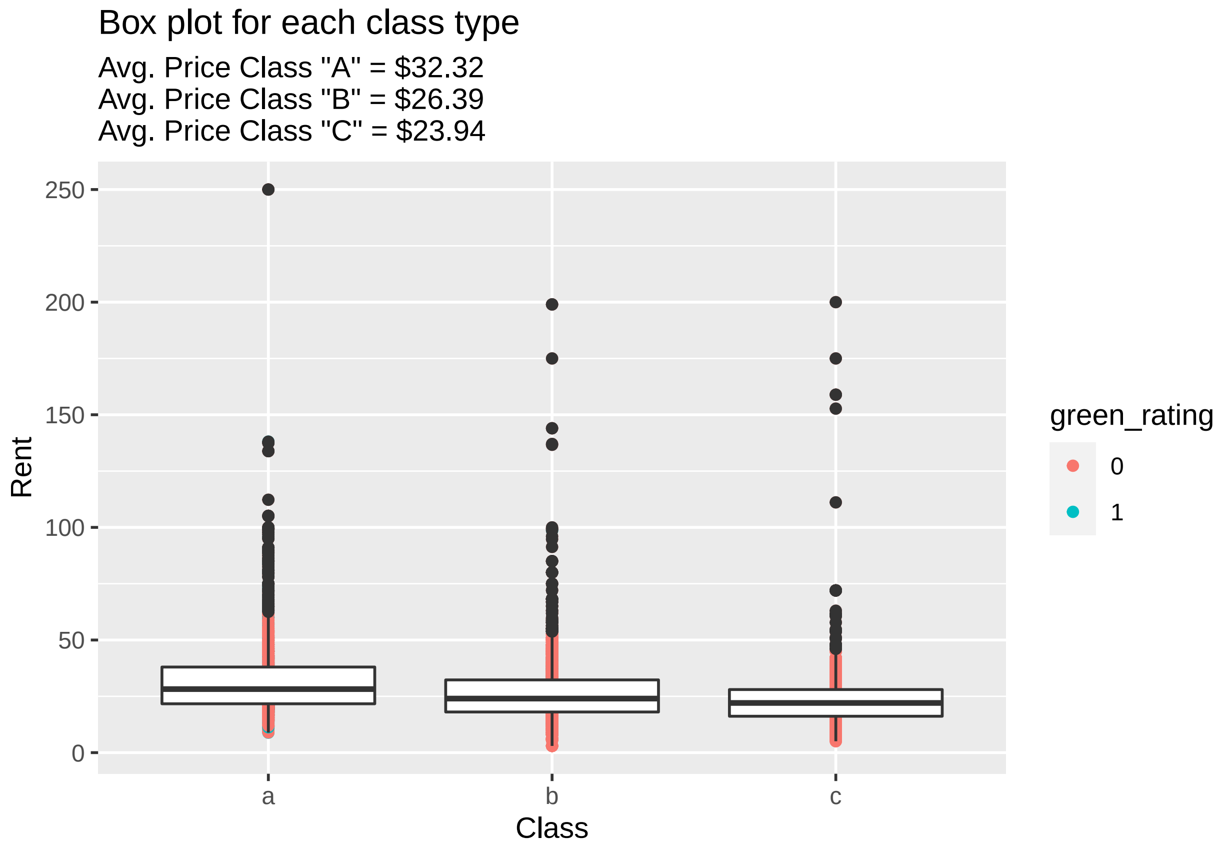



With age, we can see that there is a gradual decrease in rent as the building ages, but there are many outliers that charge very high rent rates.

Since there is a negative correlation between age and rent prices, and green buildings have a much lower average age compared to the overall average, we inferred that this can leads to generally higher rent prices for green buildings.

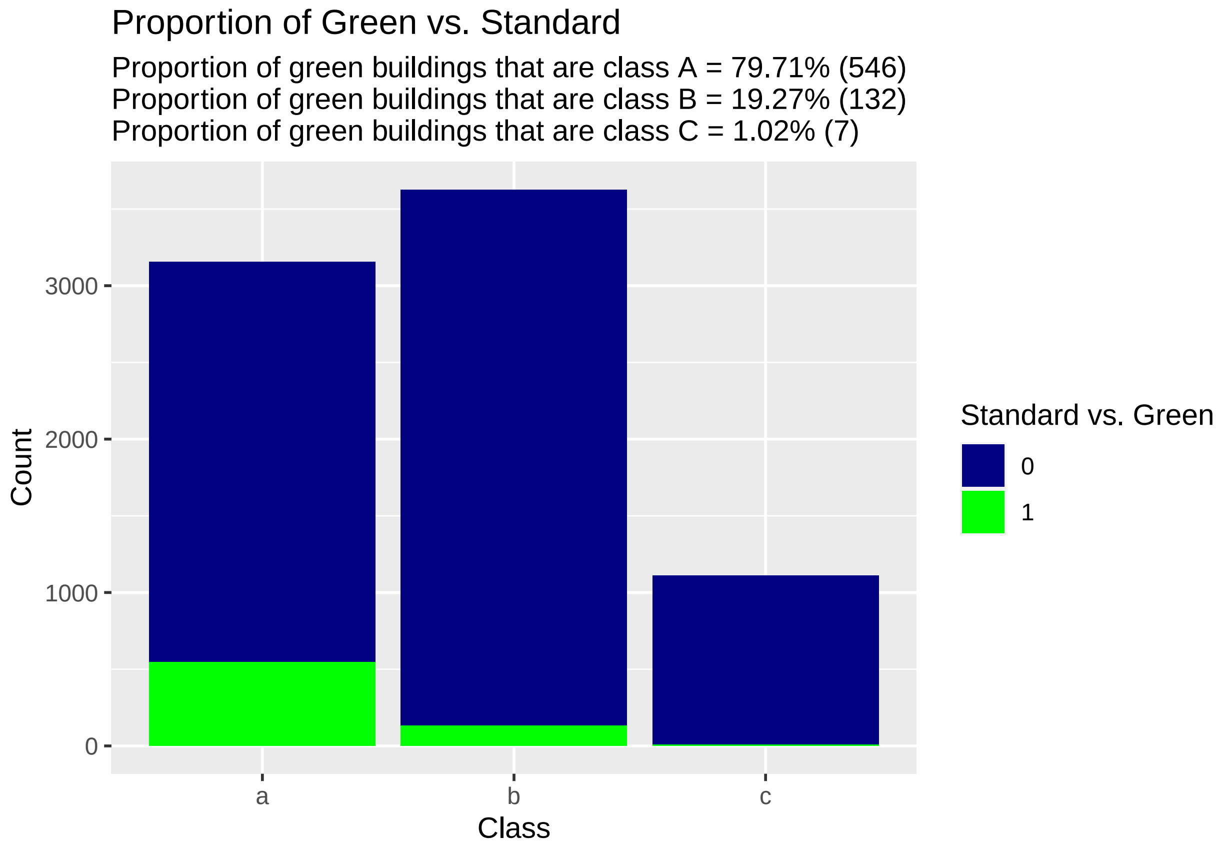
This factor coupled with their higher than average sizes supports the assumption that buildings with green status can charge a premium in rent.

# 4




After seeing how Rent changes as a function of numeric variables age and size, we wanted to see how class affects rent. Of course, the highest class building, class A, has the highest median and range, followed by B then C.

**5**




Here we can see that out of the three classes, green has it's highest number in class A. As we saw above, class A is also the most expensive classification of building. Based on green buildings being nearly 80% class A, one can claim that there is a positive correlation between green status and higher rent prices; if someone were to randomly pick a green building, 4/5 times they would choose a class A, the highest price class.

```
## Warning: Removed 74 rows containing missing values (geom_point).
```


We were curious on how Rent prices fluctuated, given the year-over-year employment growth rate that each building's region was seeing. While there is a very slight increase in the range of rent prices from the subsets 15% to 60%, we decided that this data is inconclusive. The majority grouping of buildings in the 0-7.5% range has such a wide range of rent prices that it appears that growth rate does not affect rent prices charged.

## Conclusion

From the three variables tested that provided substantial evidence (Size, Age, and Class), they indirectly showed that people paid a premium for green buildings. While this does support the stats guru's conclusion that the company could price the rent of a green building at a premium, we do not believe that these conclusions provide enough evidence to concretely say that any green building can charge a premium. What we did conclude is that the company should focus more on is the overall size of the building and developing one that would be of class A.

Size - Since buildings that are larger generally charge higher prices on average, the company should research developing a building larger than average (234,638 sqft.). Class - Buildings that are of class A, on average charge the highest rent prices at $32.32/sqft., while class B and class C charge $26.39/sqft.and $23.94/sqft., respectively.

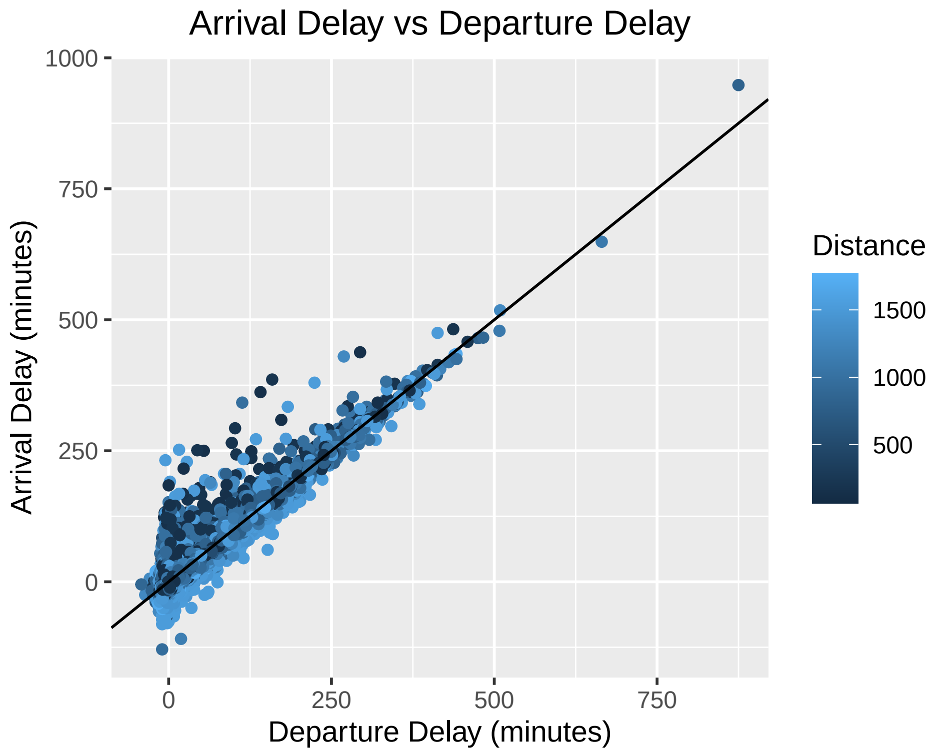
While we cannot say with certainty that a building that has a received a green rating equates to being able to charge a premium over non-green rated buildings, we do believe that a further in-depth analysis on other variables should be taken generate this answer.

## Question 2 - ABIA

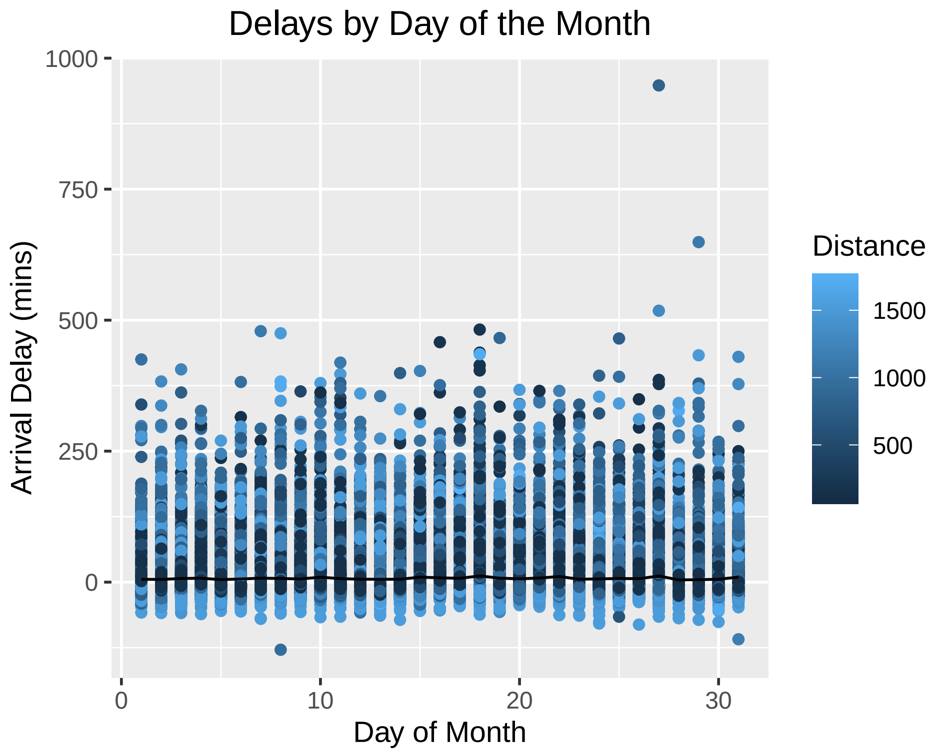
## Flights at ABIA

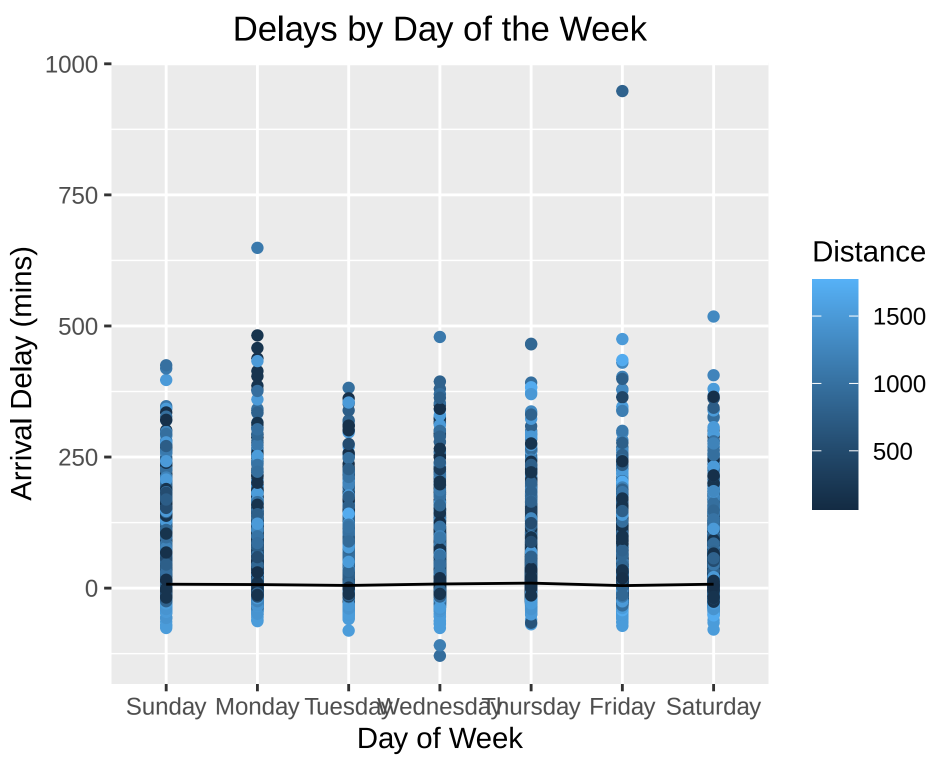
*Visual Story Telling: Create a set of related figures that tell an interesting story about flights into and out of Austin.*

**Given information about Arrival and Departure Delays, what insights can we gather?**

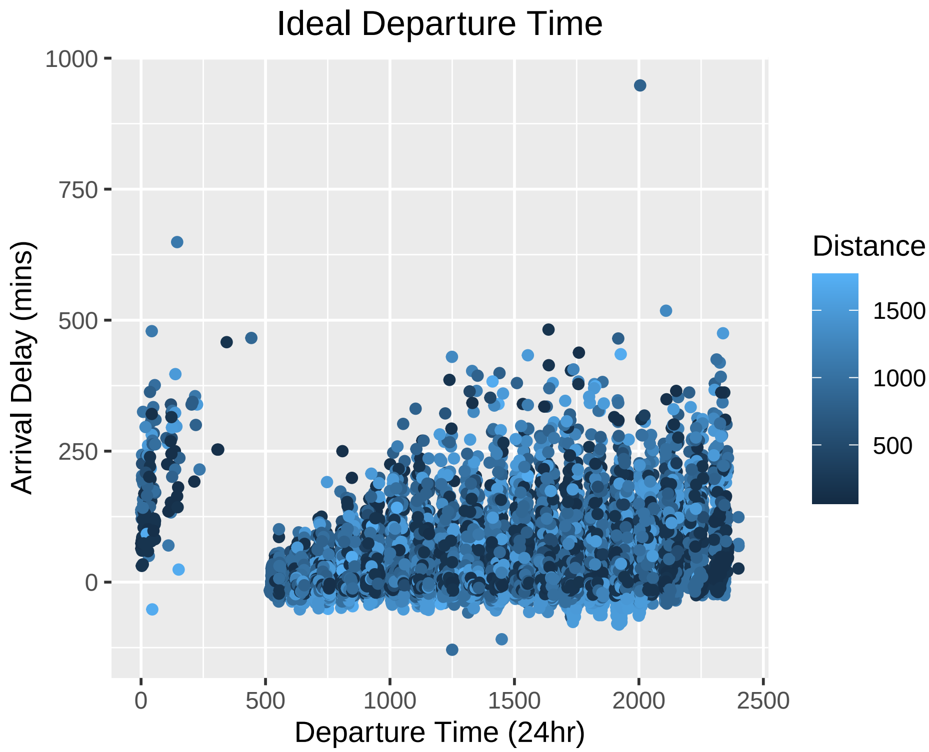



First, we plot the relationship between departure and arrival delays. This result shows, not surprisingly, that there is a relationship of approximately y=x. Arrival delays have a tendency to be higher, suggesting in-air delays, such as circling for an available runway on approach. Because arrival delays tend to have a greater influence on flying experience (particularly for connecting flights), we will use arrival delays in our analysis.

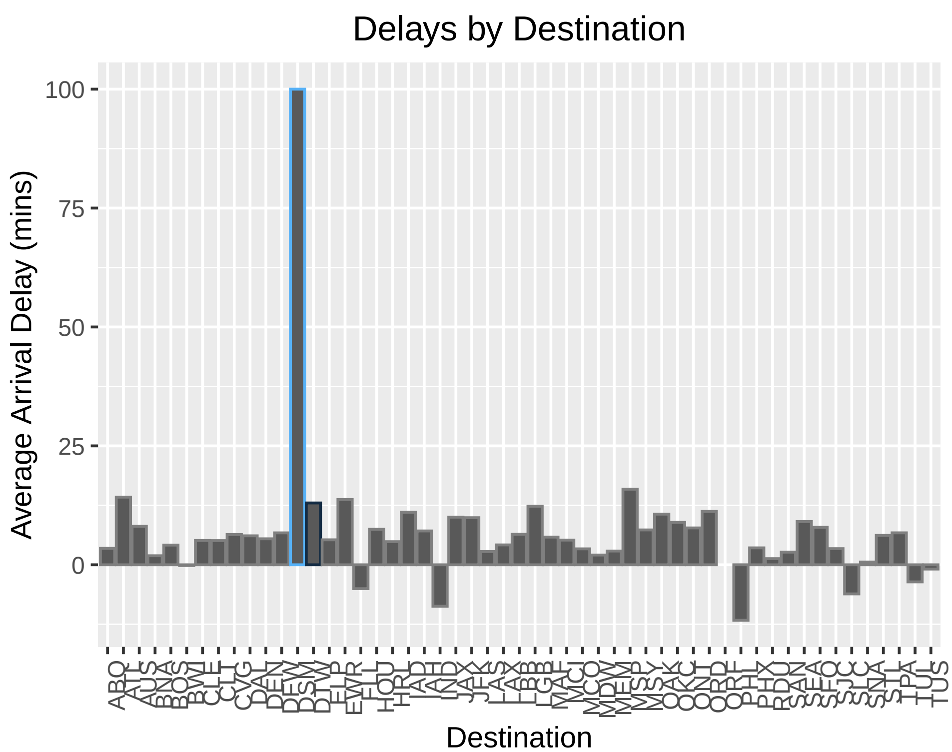



Are there any days you should avoid flying? Arrival delay is plotted against day of the month and day of the week. While there are some slight deviations, the average delay is negligible in either plot (with an overall average of 7.06 mins). Just three days of the month have an average delay over 10 minutes: the 18th, 22nd and 27th. Thursday has the highest average delay at 9.55 minutes.

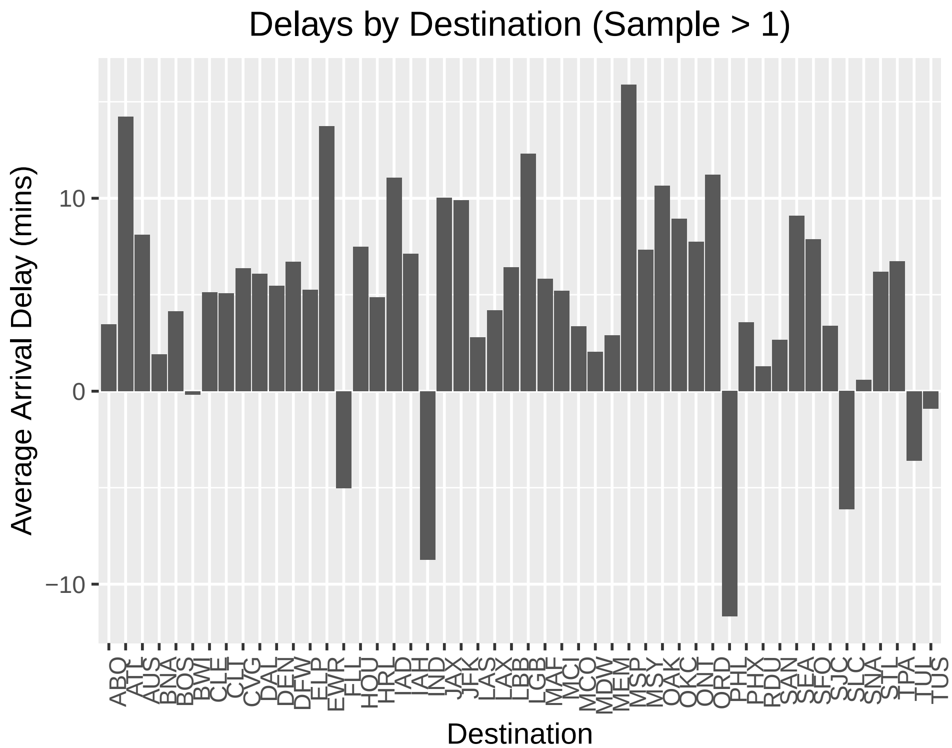



Next, we look at the best time of day to fly. Intuitively, as the day goes on, delays appear to grow. This follows a common idea that delayed flights in the morning have a domino effect on flights throughout the day. The best time to schedule a flight to avoid a delay is between 5am and 8am.

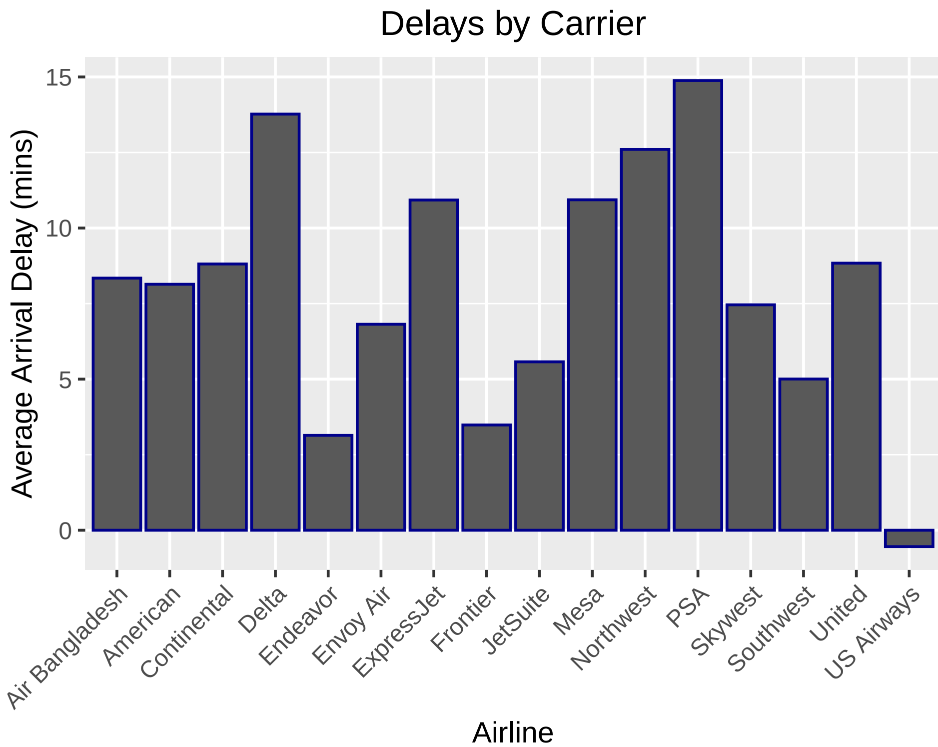



Are there any airports to avoid? Based on average arrival delay, it seems that we should never fly to Des

Moines (DSM) as it has the highest average delay by a wide margin. However, upon further inspection, there is only data from a single flight from DSM in the set.

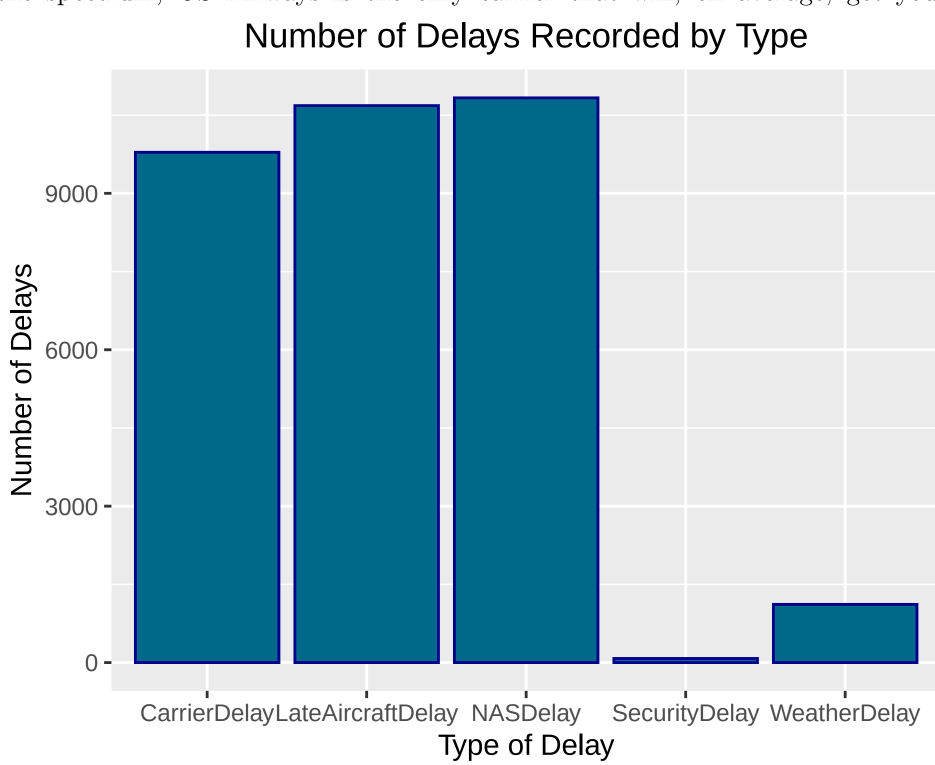



Excluding destinations with one flight (DSM, DTW and ORF), New Orleans (MSP) is now the airport with the longest average delay of over 15 minutes. In contrast, flights going into Baltimore (BWI), Fort Lauderdale (FLL), Indianapolis (IND), Philadelphia (PHL), Salt Lake City (SLC), Tulsa (TUL), and Tucson (TUS) all actually arrive ahead of schedule, on average. These are the best airports to fly into if leaving Austin and




looking to avoid a delay.

Which carriers are most likely to be delayed? Based on average arrival delay, PSA Airlines is the most likely to be responsible for your missed connection, followed by Delta. On the other end of the spectrum, US Airways is the only carrier that will, on average, get you there ahead of schedule.




Finally, what's causing all these delays anyway? Clearly, Carrier Delays, Late Aircraft, and NAS (National

Aviation System) Delays are the most significant culprits, with Security and Weather Delays relatively infrequent. This can be misleading, however, as NAS delays include all *non-extreme* weather issues, which can be reduced with corrective action by airports or the FAA, among other factors like heavy air traffic and airport operations issues. Weather Delays only refer to rare extreme weather events that cannot be mitigated.

# Question 3 - Portfolio Modeling

## Selecting the ETFS:

**In constructing our portfolio of ETF's we wanted to select a subgroup that could itself be arranged into separate portfolios of differing characteristics, but still be sufficiently diversified. Our first selection was EMB, J.P. Morgan's emerging markets bond ETF, chosen due to recent data showing emerging market bond's outperforming US and European bond ETFs. We then chose IEF, an ETF tracking 7-10 year treasury bonds, generally seen as a relatively safe, albeit underperforming, fixed income group. DVY, the iShares Select Dividend ETF, was chosen to provide our portfolio some consistent income. VIXY, a short-term volatility index, was chosen to take advantage of the current levels of volatility. XLF, the Financial Sector Select SPDR fund, was chosen to take advantage of the returns that financial services tend to provide, but have provided in a surprising way during the COVID recession. XLV is a Health Care index and was chosen as somewhat of a bet that health care related firms would be receiving some level of stimulus and a lot of attention, both from customers and investors, as treatments for COVID and distribution of PPE is widely sought after. Chosing VGK, the Vanguard FTSE Europe ETF, was due in large part to the recent stimulus package approved by EU nations that should bolster the gradual reopening we are seeing in Europe that should lead to some positive returns over the coming quarters for EU based firms. Furthermore, gaining exposure to Europe, the only major market so far neglected in the portfolio provides some naive diversification at the very least. XLP is a consumer staples ETF that should see some positive returns, and has had some very positive returns over the last 6 months, as customers are rushing to stock up on staple goods. QLD is a leveraged equity ETF that tracks the Nasdaq-100 index and seeks to return 2x the returns of the its underlying index. Given the recent returns of tech companies, which make up a considerable portion of the Nasdaq, which should continue as the US remains relatively locked down, this seemed like a very safe investment. Finally, the IWN, which tracks the Russell 2000, was chosen as a means of gaining exposure to small-cap value stocks, securities that have seen generally positive returns over growth stocks, and idea that serves as the basis for the Fama-French valuation model.**

## Selecting the First Portfolio

**While a three security portfolio is generally never a great idea, due to lack of diversification, we decided to construct one anyways to at least prove in subsequent iterations why more securities in a portfolio will generally provide better, or less volatile, returns.**

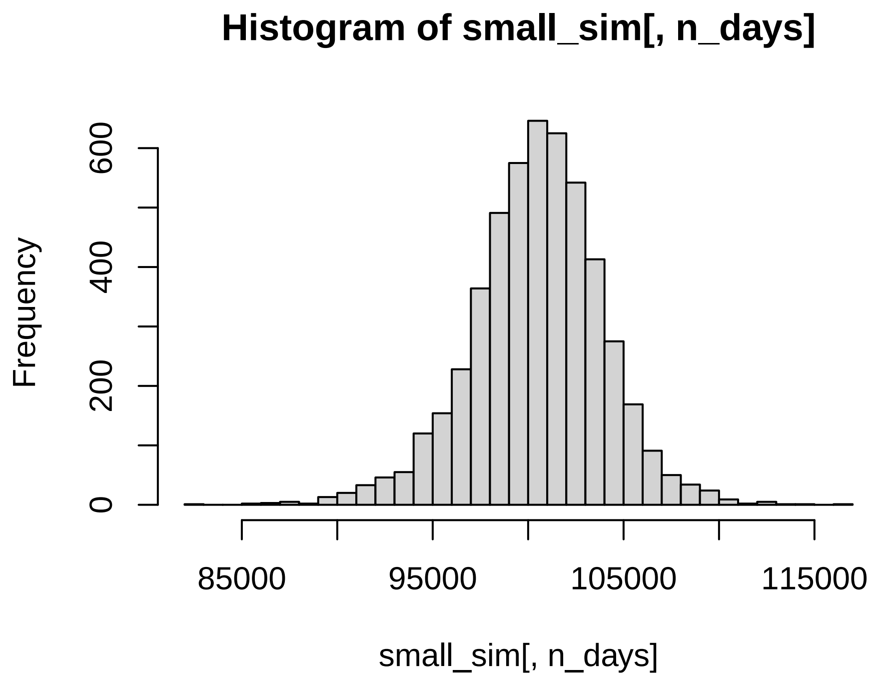
**The portfolio selected included our IEF, DVY, and VGK ETFs, which represent our safest selections. Fixed income and dividends, mixed with a European equity ETF, should provide relatively stable returns without introducing much risk. The first step was binding the returns, setting equal weights, then running the bootstrap to simulate returns over a 4 week period.**

```
#Bind first row into return matrix
small_returns = cbind(ClCl(IEFa), ClCl(DVYa), ClCl(VGKa))

small_returns = as.matrix(na.omit(small_returns))
N = nrow(small_returns)

initial_wealth = 100000
my_weights1 = c(1/3,1/3,1/3)
```

```r
set.seed(1234)

small_sim = foreach(i=1:5000, .combine='rbind') %do% {
  total_wealth = initial_wealth
  weights = my_weights1
  holdings = weights*total_wealth
  n_days = 20
  wealthtracker = rep(0,n_days)
  for(today in 1:n_days) {
    return.today = resample(small_returns,1,orig.ids=FALSE)
    holdings = weights*total_wealth
    holdings = holdings*(1+return.today)
    total_wealth = sum(holdings)
    holdings = weights*total_wealth
    wealthtracker[today] = total_wealth
  }
  wealthtracker
}

head(small_sim)
```

```
##               [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## result.1 100438.18 100490.12 100996.44 100427.89 101013.78 101103.59 101454.88
## result.2 100360.16 100128.85 101086.09 100542.25 100404.66 101207.00 101847.25
## result.3 100191.64 100991.57 101159.88 101134.93  98031.93  98890.49  98901.91
## result.4  99931.11  99195.55  98483.17  98576.38  97274.35  97907.31  98391.71
## result.5  99622.79  99617.88  99831.05  99821.26  99552.43  99885.02  99916.33
## result.6  99995.07 100588.57 100878.57 100179.44 100044.65 100194.19 100276.95
##               [,8]      [,9]     [,10]     [,11]     [,12]     [,13]     [,14]
## result.1 101572.10 101598.19 101932.91 101047.46 100800.59 100989.12 101263.44
## result.2 102063.81 101660.37 102683.81 102547.35 102219.66 102215.98 102115.85
## result.3  99226.16  99315.28  99208.18  99074.70 100147.48 100543.86  93227.99
## result.4  98063.86  98274.92  98458.72  98588.95  98570.32  98208.56  98088.66
## result.5 100158.64  99859.18 100054.72  99793.72  99546.13  99354.10  99650.15
## result.6 100392.82 100741.67 101043.08 100505.67 100518.24 100689.74  99185.86
##              [,15]     [,16]     [,17]     [,18]     [,19]     [,20]
## result.1 101271.19 101955.61 101177.86 101522.17 101742.03 101550.51
## result.2 102328.49 103034.49 104870.20 104816.33 104285.33 104011.75
## result.3  93699.71  93432.98  93841.14  93733.26  92809.40  93270.83
## result.4  98332.50  98473.20  97790.74  97278.41  96920.74  96916.27
## result.5  98445.24  97815.75  97668.36  97757.64  97655.48  97421.82
## result.6  98017.46  98806.53  99159.30  99093.34 100622.99 100668.34
```

```r
hist(small_sim[,n_days], 25)
```



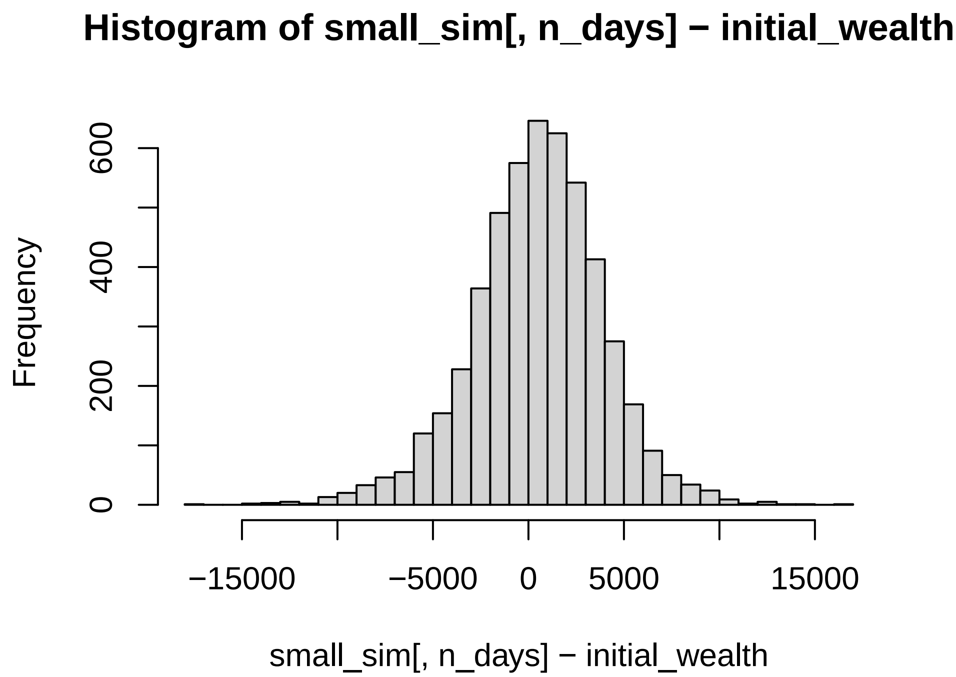

```r
mean(small_sim[,n_days])
```

```
## [1] 100475.4
```

```r
mean(small_sim[,n_days] - initial_wealth)
```

```
## [1] 475.3521
```

```r
hist(small_sim[,n_days]- initial_wealth, breaks=30)
```




```r
# 5% value at risk:
quantile(small_sim[,n_days]- initial_wealth, prob=0.05)
```

```
##        5%
## -5341.957
```

Despite our best efforts to reduce volatility, it is clear in our results that we were not successful. Our mean return over the 20 day period was $422.21. Not surprising given the fact that investments in fixed income and dividend stocks are generally not made for huge returns on equity, but rather for the provision of income and their relative safety in times of uncertainty.

It should be noted, however, that we are in a period of unprecedented uncertainty and a time in which treasury securities' returns are historically low. Given these two facts, it is equally unsurprising that our VaR, or Value at Risk, a measure that estimates how much a set of investments might lose at the 5% confidence interval, is -$5,277.57. That is considerably more than our mean returns which shows that this is likely not a great portfolio to invest in at this time. Even though this is the least likely scenario, it is a measure that some investors put a lot of weight in given the relative increase in "black swan" events.

It should be noted of course that our distribution is centered around slightly positive returns, which means on average an investor can expect positive returns from this portfolio.

## Choosing the Second Portfolio

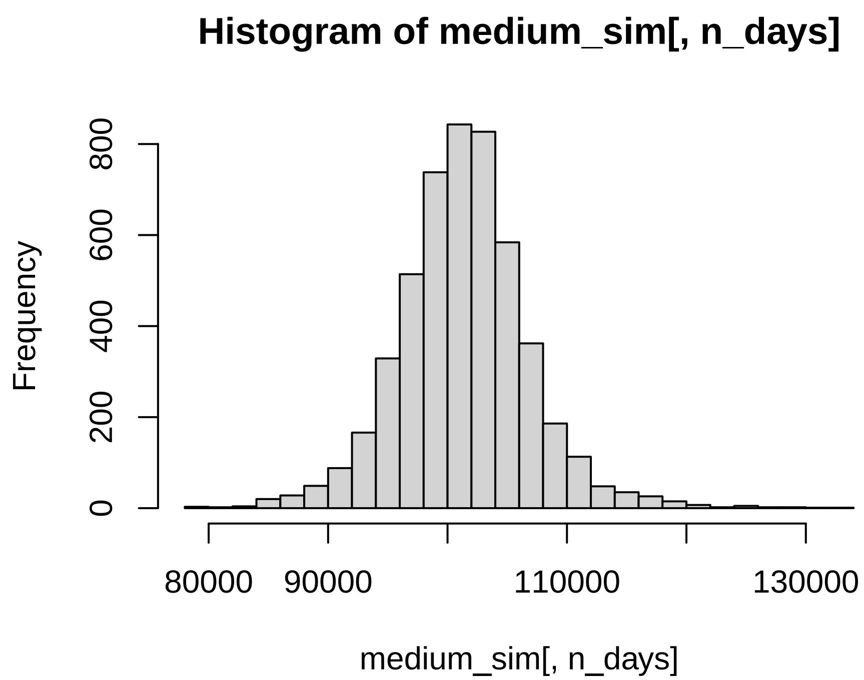
In choosing the second portfolio, we sought to build on the first rather than select a whole new subset of securities. We did this to stay consistent with our effort to show the effect of diversification in a portfolio. For this portfolio, we added three more securities, EMB, XLF, and QLD. We saw these additions as relatively aggressive additions because of the outsized returns that we expect these securities to provide in the current market conditions.

```r
medium_returns = cbind(ClCl(IEFa), ClCl(DVYa), ClCl(VGKa), ClCl(EMBa),
                      ClCl(XLFa), ClCl(QLDa))

medium_returns = as.matrix(na.omit(medium_returns))
N = nrow(medium_returns)

initial_wealth = 100000
my_weights2 = c(1/6,1/6,1/6)

set.seed(1234)

medium_sim = foreach(i=1:5000, .combine='rbind') %do% {
  total_wealth = initial_wealth
  weights = my_weights2
  holdings = weights*total_wealth
  n_days = 20
  wealthtracker = rep(0,n_days)
  for(today in 1:n_days) {
    return.today = resample(medium_returns,1,orig.ids=FALSE)
    holdings = weights*total_wealth
    holdings = holdings*(1+return.today)
    total_wealth = sum(holdings)
    holdings = weights*total_wealth
    wealthtracker[today] = total_wealth
  }
  wealthtracker
}

head(medium_sim)
```

```
##                [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## result.1 100850.29 100978.28 102116.52 101305.41 102077.00 102519.03 102892.80
```

```
## result.2 100858.26 100484.63 101901.11 101154.42 101425.21 102841.46 104124.81
## result.3 100374.45 100842.59 100999.22 101271.86  96854.84  96699.92  96858.23
## result.4 100330.93  99327.96  98409.35  98256.40  96856.69  97674.74  98016.43
## result.5  99184.49  99201.46  99293.42  99383.69  99152.69  99615.76  99688.18
## result.6 100017.10 100883.99 101593.37 100525.27 100325.84 100216.68 100682.66
##               [,8]      [,9]     [,10]     [,11]     [,12]     [,13]     [,14]
## result.1 103018.23 103011.87 103916.09 102593.90 101738.74 102026.27 102593.16
## result.2 104271.15 103846.87 105530.39 105481.08 104844.56 105029.33 104722.20
## result.3  96965.17  97077.32  96720.72  96528.84  98316.99  98437.68  89237.62
## result.4  97738.31  98686.57  98965.48  99200.17  99422.11  99362.97  99531.90
## result.5  99587.35  99248.19  99436.88  99203.38  98555.59  98604.48  98566.14
## result.6 100805.40 100960.95 101384.94 101173.91 101048.83 101280.88  98424.83
##              [,15]     [,16]     [,17]     [,18]     [,19]     [,20]
## result.1 102654.96 103237.23 102274.22 102966.53 103266.27 103131.31
## result.2 104994.94 105951.37 107851.89 108307.62 107479.16 107110.46
## result.3  89893.93  89517.27  90184.42  90413.30  88739.21  89362.95
## result.4  99970.25 100129.70  99076.99  98685.78  98473.30  98130.60
## result.5  96899.66  95792.26  95511.20  95877.52  95719.85  95476.78
## result.6  96374.63  97694.01  98248.22  98374.32  99961.87 100232.91
```

```r
hist(medium_sim[,n_days], 25)
```



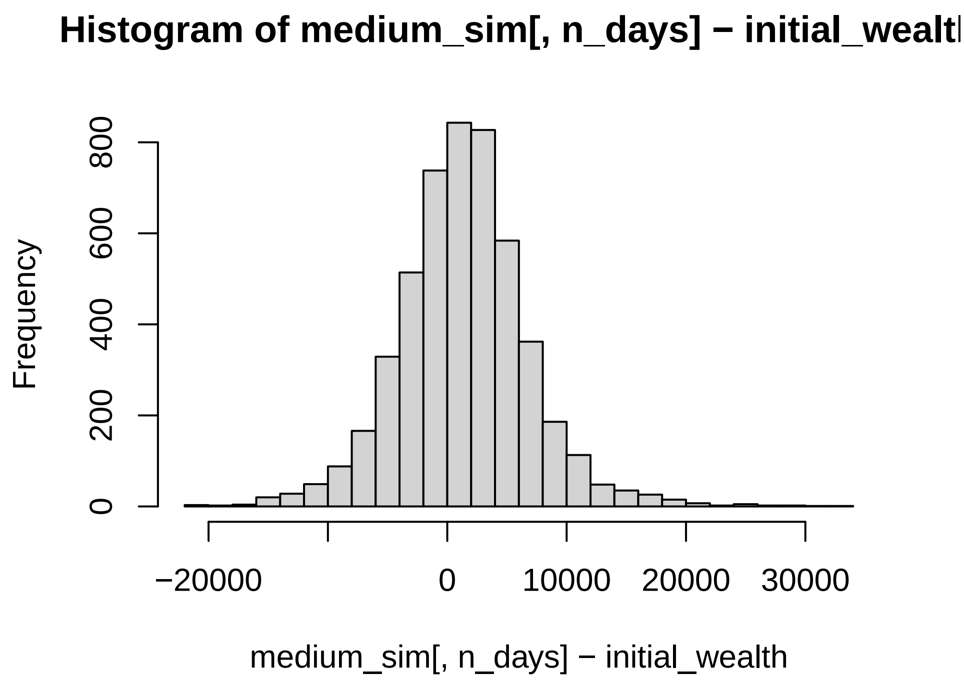

```r
mean(medium_sim[,n_days])
```

```
## [1] 101356.8
```

```r
mean(medium_sim[,n_days] - initial_wealth)
```

```
## [1] 1356.775
```

```r
hist(medium_sim[,n_days]- initial_wealth, breaks=30)
```




```r
# 5% value at risk:
quantile(medium_sim[,n_days]- initial_wealth, prob=0.05)
```

```
##        5%
## -7085.436
```

**As we suspected, the returns did exceed the safe portfolio created in the first iteration. In fact, the mean returns, $1318.80 far exceeded the returns of the initial portfolio.**

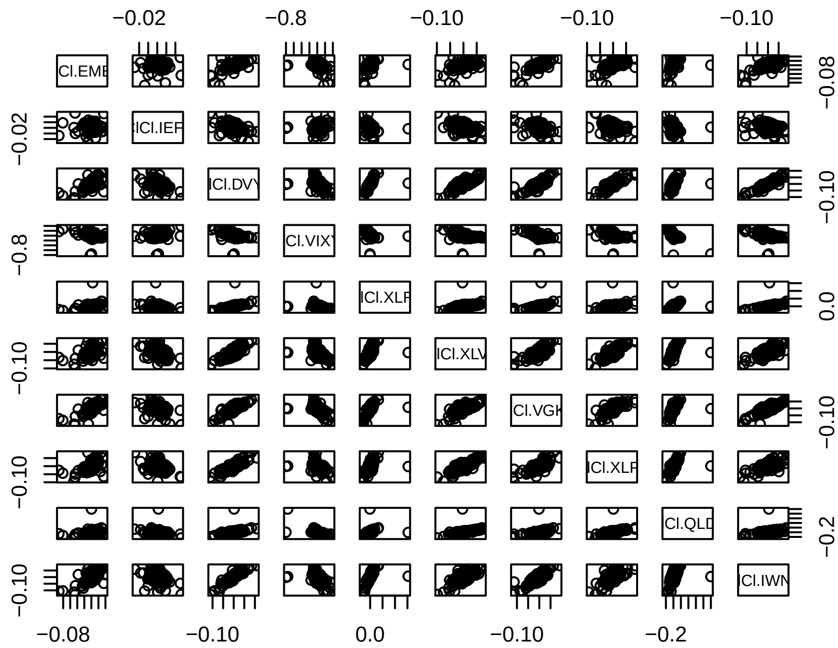
**However, the VaR of this portfolio, -$7398.02 also exceeds the first portfolio. Considering the aggressive additions for this portfolio, that is not completely surprising, but it doesn't back up our hypothesis that additional securities, even aggressive securities, would reduce the overall risk of the portfolio. It should be noted of course that our distribution is centered around slightly positive returns, which means on average an investor can expect positive returns from this portfolio.**

##Selecting the Final Portfolio

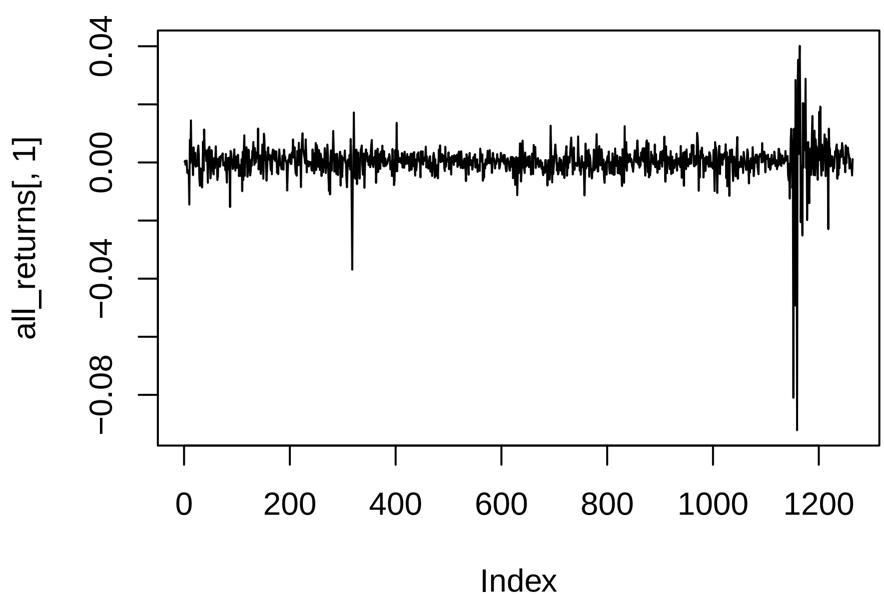
**For the final portfolio, we added all of the rest of our securities.**

```r
#Bind first row into return matrix
all_returns = cbind(ClCl(EMBa), ClCl(IEFa), ClCl(DVYa), ClCl(VIXYa), ClCl(XLFa),
                    ClCl(XLVa), ClCl(VGKa), ClCl(XLPa), ClCl(QLDa),  ClCl(IWNa))

#Get rid of NA first row
all_returns = as.matrix(na.omit(all_returns))
N = nrow(all_returns)
```

```r
#plots pairs and historical returns
pairs(all_returns)
```



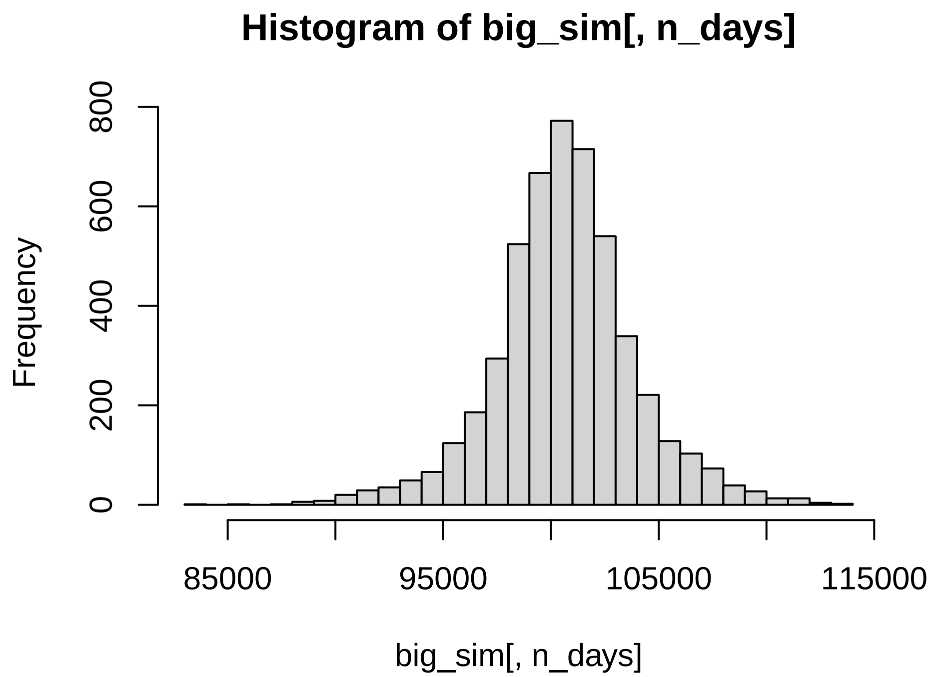

```r
plot(all_returns[,1],type='l')
```



```r
initial_wealth = 100000
my_weights3 = c(.1,.1,.1,.1,.1,.1,.1,.1,.1,.1)

set.seed(1234)

big_sim = foreach(i=1:5000, .combine='rbind') %do% {
  total_wealth = initial_wealth
  weights = my_weights3
  holdings = my_weights3*total_wealth
  n_days = 20
  wealthtracker = rep(0,n_days)
  for(today in 1:n_days) {
    return.today = resample(all_returns,1,orig.ids=FALSE)
    holdings = weights*total_wealth
    holdings = holdings*(1+return.today)
    total_wealth = sum(holdings)
    holdings = weights*total_wealth
    wealthtracker[today] = total_wealth
  }
  wealthtracker
}

head(big_sim)
```

```
##               [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## result.1 100584.51 100522.67 100853.37 100329.40 100713.93 100881.77 101053.60
## result.2 100299.75  99962.32 100476.47 100198.70 100320.72 100251.82 100677.58
## result.3 100009.31 100751.23 100615.56 100841.19  97841.32  98034.57  98115.57
## result.4  99847.39  99610.87  99597.89  99283.37  99328.66  99762.15  99490.62
## result.5  99564.00  99387.72  99232.71  99309.53  99107.16  99225.45  99317.19
## result.6  99822.94 100160.97 100513.66 100017.05  99740.71  99475.58  99572.79
##               [,8]      [,9]     [,10]     [,11]     [,12]     [,13]     [,14]
## result.1 100961.94 100994.32 101670.74 101519.87 101352.28 101623.60 101950.60
## result.2 100654.12 100348.18 101258.42 101307.92 100904.08 100884.46 100934.76
## result.3  98431.35  98398.19  98276.85  98005.31  99001.46  99046.83  93077.88
## result.4  99522.46  99985.71 100253.37 100324.26 100385.43 100254.14 100537.52
## result.5  99374.69  99320.76  99463.03  99248.31  99303.94  98948.46  98957.64
## result.6  99482.47  99555.36  99868.58  99888.97 100091.95 100174.60  98751.03
##              [,15]     [,16]     [,17]     [,18]     [,19]     [,20]
## result.1 102038.61 102223.12 101902.87 102269.54 102585.61 102597.70
## result.2 100873.63 101319.69 102741.73 102771.76 102870.73 102775.48
## result.3  93357.42  93238.43  93356.13  93279.14  92495.69  92888.64
## result.4 100509.70 100422.06  99925.90  99674.65  99401.97  99522.90
## result.5  98080.06  98010.85  98132.28  98301.25  98234.14  98161.05
## result.6  98228.29  98858.54  99330.65  99526.06 100401.72 100416.04
```

```r
hist(big_sim[,n_days], 25)
```



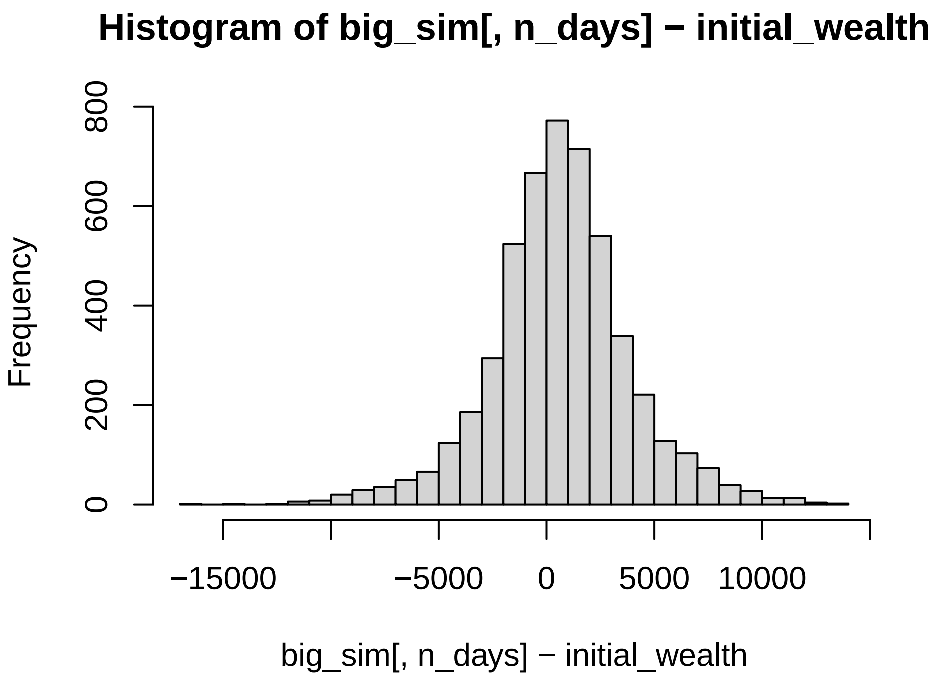

```
mean(big_sim[,n_days])
```

```
## [1] 100638.4
```

```
mean(big_sim[,n_days] - initial_wealth)
```

```
## [1] 638.4383
```

```
hist(big_sim[,n_days]- initial_wealth, breaks=30)
```




```
# 5% value at risk:
quantile(big_sim[,n_days]- initial_wealth, prob=0.05)
```

```
##        5%
## -4612.515
```

**The results of our final portfolio provide the most interesting information. One, or several, of the securities added to our six security portfolio essentially erased some of the returns from our six security portfolio. On the other hand, our VaR was also reduced significantly. It should be noted of course that our distribution is centered around slightly positive returns, which means on average an investor can expect positive returns from this portfolio.**

**In fact, this portfolio created returns above the initial 3 stock portfolio, with risk less than the 6 stock portfolio. This generally follows the theory that higher returns result in higher risk. However, the higher risk associated with adding 7 additional securities to the intial portfolio is not significantly higher, which may show that we successfully diversified while increasing returns. This is certainly not conclusive, but a strong step towards creating an effective portfolio of ETFs.**

## Question 4 - Market Segmentation

NutrientH20 has requested a market segment analysis on their social media audience. We are given 7,882 twitter followers along with the frequency with which they've tweeted about 36 different topics.

From a basic sweep over the various topics, one might assume that some are more correlated than others. For instance, those tweeting about Health and Nutrition are likely to be the same people tweeting about personal fitness. However, given the multitude of possible groupings of these topics, we will want to automate the process. For that, we conduct a k-means clustering analysis. While most of the topics are relevant, we have removed the "chatter" and "uncategorized" topics from our topic list as they appear to have been used by annotators to classify tweets that didn't fit into other categories. They likely represent a multitude of topics so we omit them to avoid the k-means model assuming that they represent one topic.

To choose the optimal number of clusters, we first look at the Gap Statistic for various values of k.

```
## Warning: did not converge in 10 iterations
## Warning: Quick-TRANSfer stage steps exceeded maximum (= 394100)
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
## Warning: Quick-TRANSfer stage steps exceeded maximum (= 394100)
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
## Warning: Quick-TRANSfer stage steps exceeded maximum (= 394100)

## Warning: Quick-TRANSfer stage steps exceeded maximum (= 394100)
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
## Warning: Quick-TRANSfer stage steps exceeded maximum (= 394100)
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
## Warning: Quick-TRANSfer stage steps exceeded maximum (= 394100)
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
## Warning: Quick-TRANSfer stage steps exceeded maximum (= 394100)
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
## Warning: Quick-TRANSfer stage steps exceeded maximum (= 394100)
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
## Warning: Quick-TRANSfer stage steps exceeded maximum (= 394100)
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
## Warning: Quick-TRANSfer stage steps exceeded maximum (= 394100)
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
## Warning: Quick-TRANSfer stage steps exceeded maximum (= 394100)
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations
## Warning: Quick-TRANSfer stage steps exceeded maximum (= 394100)
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
```

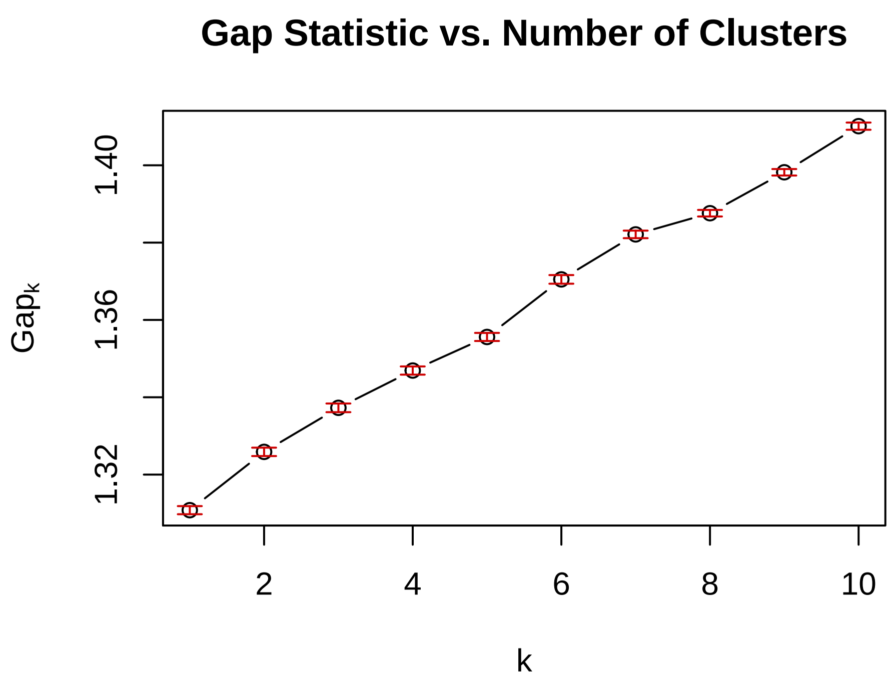
```
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
## Warning: Quick-TRANSfer stage steps exceeded maximum (= 394100)

## Warning: Quick-TRANSfer stage steps exceeded maximum (= 394100)
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
## Warning: Quick-TRANSfer stage steps exceeded maximum (= 394100)
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
```

```
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
```




With no clearn kinks in this gap statistic plot, we are not getting a clear definition of what number of clusters to choose, and the data is not converging cleanly before k =10. However, we decide to choose 5 clusters to hopefully identify easily interpretable market segments within the NutrientH20 social media following.

Now we can construct our k-means clustering with k=5 and 25 initial configurations. The k-means algorithm clusters by the following percentages (we will look at the characteristics of these clusters shortly):

Group 1 - 61.38% of twitter following

Group 2 - 9.99% of twitter following

Group 3 - 9.07% of twitter following

Group 4 - 7.66% of twitter following

Group 5 - 11.89% of twitter following

After automating the market segment split via kmeans, we confirm our cluster choices by conducting a k-means++ initialization and comparing the centroids of our clusters from each initialization method.

Within-cluster average distances for the first run are identical for k-means and k-means++ after 25 iterations.

Kmeans average distances:

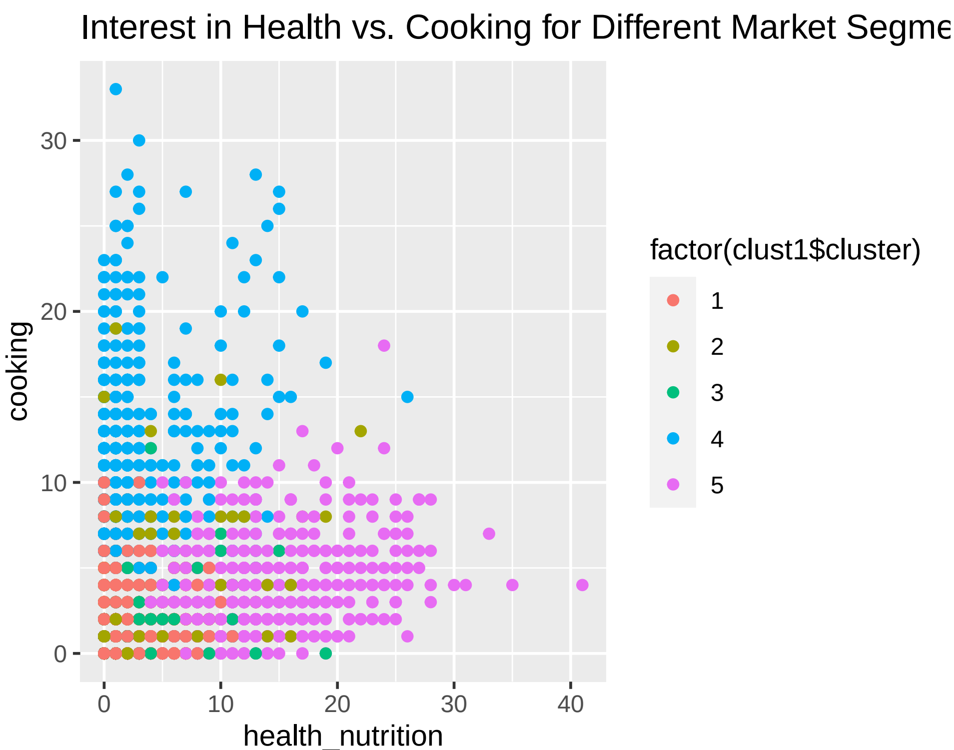
```
## [1] 23839.26 27831.82 30402.61 30656.29 96126.34
```

K-means++ average distances:

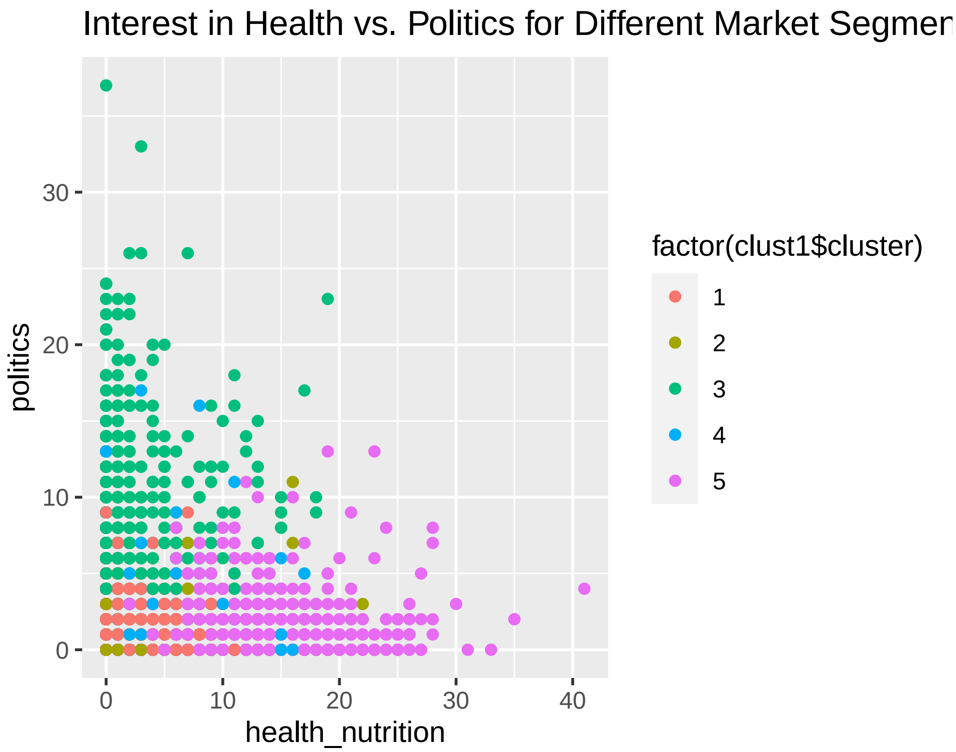
```
## [1] 23839.26 27831.82 30402.61 30656.29 96126.34
```

We can see that from both k-means and k-means++ initialization, we receive identical groups for our 5 clusters.

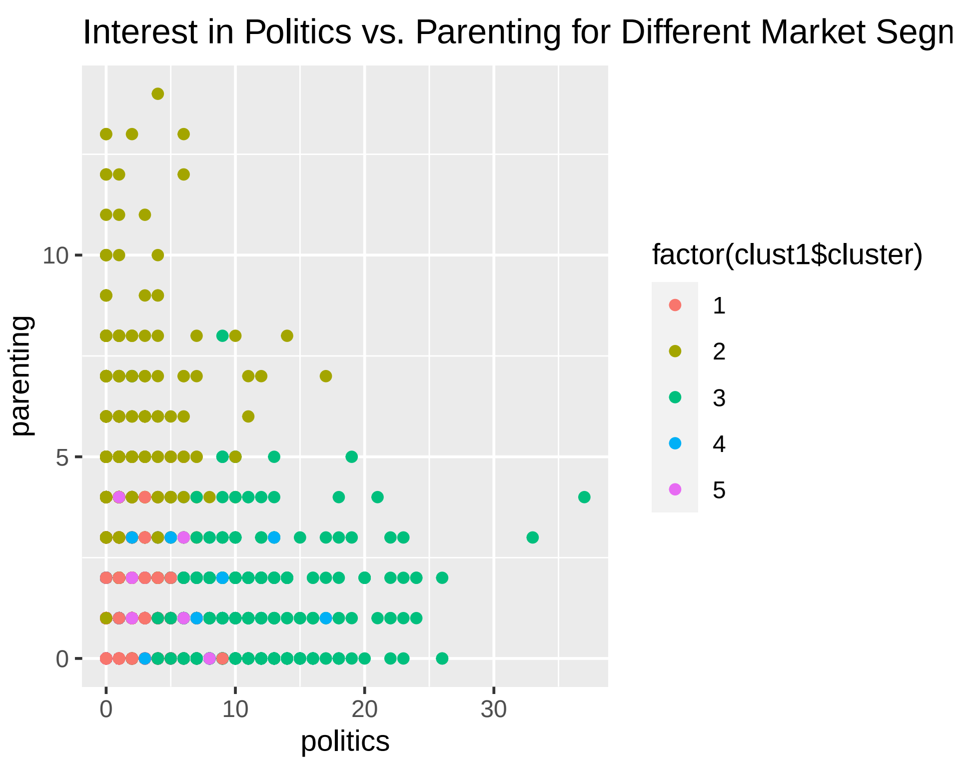
Having established five different market segments, we can explore the characteristics of each segment visually through a few topic comparison plots.



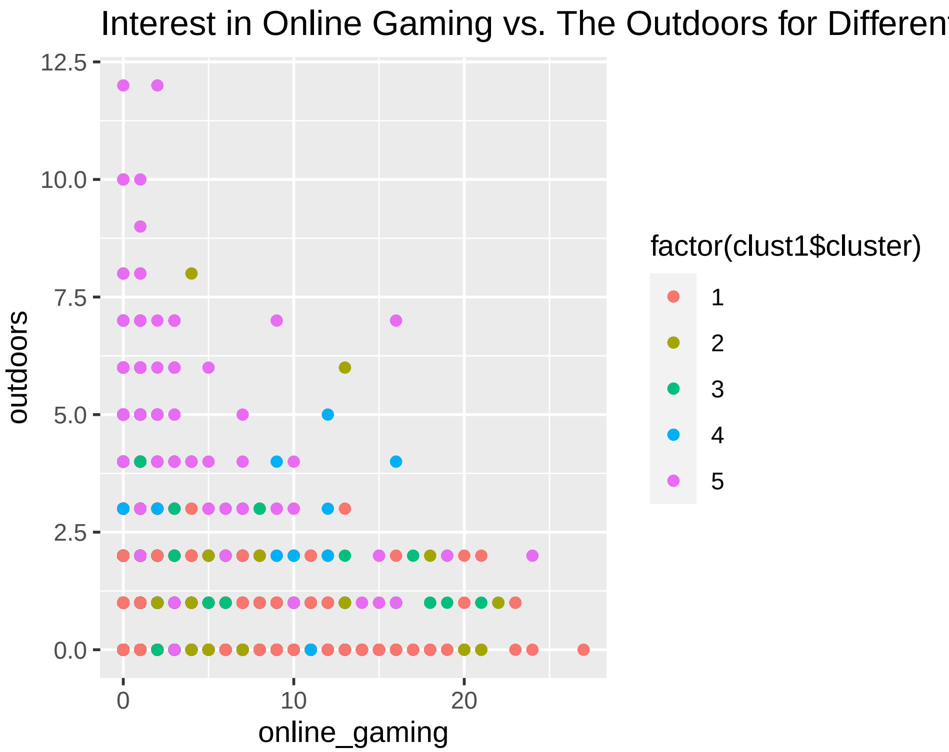

Here we see an interesting split between clusters - although group 5 seems to care greatly about health and nutrition, they care less about cooking than does group 4. This gives us the inkling that group 5 might be our “fit” group.




Here we can see that group 3 concerns themselves with politics far more often than do the other twitter followers.




Here we can see that group 2 is more concerned with parenting than any other group. This might be a more family-oriented audience.




Finally, here we again see that our "fit" group - group 5- is the most interested in the outdoors, while group 1 stands out as having interests in online gaming.

Thus, we can see some distinct characteristics about each cluster of NutrientH20 twitter followers. We can look at the top 5 defining topics for the centroids of each cluster to start to develop a better idea of who falls into each group.

### "The General Public"

```
##  photo_sharing    college_uni current_events       shopping  online_gaming
##       2.311906       1.511782       1.449153       1.269740       1.168458
```

Group 1 tends to share photos and post about online gaming, college, and shopping at higher rates than other groups. However, none of the topics stand out with a particularly high importance. It appears k-means has defined this cluster as a "catch-all" for individuals who don't fit the four molds described below.

### "Family-Oriented

```
## sports_fandom      religion          food     parenting        school
##      5.864213      5.241117      4.535533      4.016497      2.687817
```

Group 2 concerns themselves with sports fandom, religion, food and family topics. We might assume this to be middle-aged parents and a generally more family-oriented audience.

### "Internationally Aware"

```
##      politics        travel          news photo_sharing     computers
##      8.823776      5.566434      5.195804      2.514685      2.467133
```

Group 3 posts frequently about politics, news, and travel. These might be internationally inclined, exploratory individuals.

### “The Insta Girls”

```
##          cooking    photo_sharing          fashion           beauty
##        10.607616         6.082781         5.480132         3.804636
## health_nutrition
##         2.271523
```

Group 4 tends to frequently share photos and concern themselves with cooking, fashion, and beauty.

### “The Fit-Focused”

```
## health_nutrition personal_fitness          cooking    photo_sharing
##        11.843116         6.354322         3.252935         2.704376
##         outdoors
##         2.672359
```

Group 5 posts are very much focused on health, nutrition, fitness, and to a lesser extent the outdoors. This may very well be NutrientH20’s target market.

Given NutrientH20’s branding, we might assume some interest in the seemingly health and fitness-focused market segment that is group 5. Below we have compiled the 937 “group 3” members that tend to be health, nutrition, and fitness inclined. We display the first five user codes as an example.

```
## [1] "hmjoe4g3k" "y2g68vhkf" "4odi1h6wq" "qg1jsawuy" "af7eqjw8r" "p94myqeo6"
```

Without full information on its marketing strategy, we cannot be certain as to which consumers NutrientH20 wishes to target. However, we suspect that the “Fit-Focused” group, as characterized above by interests in health and personal fitness, might be the most receptive of Nutrient H20’s marketing material. We should note that this demographic only makes up 12% of NutrientH20’s Twitter following, and still might find value in marketing to “The General Public” (61% of following) or “The Insta Girls” (8% of following).

We hope this has helped NutrientH20 identify its target market among its twitter followers and be able to target advertising to this group of customers more effectively.

# Question 5 - Author Attribution

# Step 1:

The first step to predicting the author of a given text is to create the train set from the C50Train set:

```r
library(tm)
```

```
## Loading required package: NLP

##
## Attaching package: 'NLP'

## The following object is masked from 'package:ggplot2':
##
##     annotate

##
## Attaching package: 'tm'

## The following object is masked from 'package:arules':
##
##     inspect

## The following object is masked from 'package:mosaic':
##
```

```
##     inspect
```

```
library(magrittr)
```

```
##
## Attaching package: 'magrittr'

## The following object is masked from 'package:purrr':
##
##     set_names

## The following object is masked from 'package:tidyr':
##
##     extract
```

```
library(slam)
library(proxy)
```

```
## Registered S3 methods overwritten by 'proxy':
##   method               from
##   print.registry_field registry
##   print.registry_entry registry

##
## Attaching package: 'proxy'

## The following object is masked from 'package:Matrix':
##
##     as.matrix

## The following objects are masked from 'package:stats':
##
##     as.dist, dist

## The following object is masked from 'package:base':
##
##     as.matrix
```

```
library(caret)
```

```
##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
##     lift

## The following object is masked from 'package:mosaic':
##
##     dotPlot
```

```
library(plyr)
```

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

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)

## ------------------------------------------------------------------------------

##
```

```
## Attaching package: 'plyr'

## The following object is masked from 'package:purrr':
##
##     compact

## The following object is masked from 'package:mosaic':
##
##     count

## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
library(dplyr)
library(ggplot2)
library('e1071')
library(randomForest)
```

```
## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':
##
##     margin

## The following object is masked from 'package:dplyr':
##
##     combine
```

```
library(ggplot2)
library(class)

train=Sys.glob('C:/Users/Christian/Documents/UT MSBA/Predictive Modelling/C50train/*')

#Creating training dataset
authortrain=NULL
labels=NULL

for (name in train)
{
  author=substring(name,first=50)
  article=Sys.glob(paste0(name,'/*.txt'))
  authortrain=append(authortrain,article)
  labels=append(labels,rep(author,length(article)))
}


readerPlain <- function(fname)
{
  readPlain(elem=list(content=readLines(fname)),
            id=fname, language='en')
}
```

```r
comb = lapply(authortrain, readerPlain)
names(comb) = authortrain
names(comb) = sub('.txt', '', names(comb))


corp_train=Corpus(VectorSource(comb))
```

## Step 2:

The next thing that we need to do is wrangle the data a bit to make the text comparable across files and remove any numbers or characters that might offset the models.

The second part of this step is to create our Document Term Matrix and then our TF-IDF matrix so that we can assess the number of occurrences of a particular word in any given document and across all documents.

```r
corp_train = tm_map(corp_train, content_transformer(tolower))
```

```
## Warning in tm_map.SimpleCorpus(corp_train, content_transformer(tolower)):
## transformation drops documents
```

```r
corp_train = tm_map(corp_train, content_transformer(removeNumbers))
```

```
## Warning in tm_map.SimpleCorpus(corp_train, content_transformer(removeNumbers)):
## transformation drops documents
```

```r
corp_train = tm_map(corp_train, content_transformer(removePunctuation))
```

```
## Warning in tm_map.SimpleCorpus(corp_train,
## content_transformer(removePunctuation)): transformation drops documents
```

```r
corp_train = tm_map(corp_train, content_transformer(stripWhitespace))
```

```
## Warning in tm_map.SimpleCorpus(corp_train,
## content_transformer(stripWhitespace)): transformation drops documents
```

```r
corp_train = tm_map(corp_train, content_transformer(removeWords),stopwords("en"))
```

```
## Warning in tm_map.SimpleCorpus(corp_train, content_transformer(removeWords), :
## transformation drops documents
```

```r
DTM_train = DocumentTermMatrix(corp_train)
DTM_train
```

```
## <<DocumentTermMatrix (documents: 2500, terms: 32572)>>
## Non-/sparse entries: 542861/80887139
## Sparsity           : 99%
## Maximal term length: 41
## Weighting          : term frequency (tf)
```

```r
DTM_tr=removeSparseTerms(DTM_train,.99)
tf_idf = weightTfIdf(DTM_tr)
DTM_train<-as.matrix(tf_idf)
tf_idf
```

```
## <<DocumentTermMatrix (documents: 2500, terms: 3395)>>
## Non-/sparse entries: 382971/8104529
## Sparsity           : 95%
## Maximal term length: 41
```

```
## Weighting          : term frequency - inverse document frequency (normalized) (tf-idf)
```

## Step 3:

Do the same with the test data:

```r
test=Sys.glob('C:/Users/Christian/Documents/UT MSBA/Predictive Modelling/C50test/*')

authortest=NULL
labels1=NULL

for (name in test)
{
  author1=substring(name,first=50)
  article1=Sys.glob(paste0(name,'/*.txt'))
  authortest=append(authortest,article1)
  labels1=append(labels1,rep(author1,length(article1)))
}


comb1 = lapply(authortest, readerPlain)
names(comb1) = authortest
names(comb1) = sub('.txt', '', names(comb1))


corp_test=Corpus(VectorSource(comb1))


corp_test = tm_map(corp_test, content_transformer(tolower))
```

```
## Warning in tm_map.SimpleCorpus(corp_test, content_transformer(tolower)):
## transformation drops documents
```

```r
corp_test = tm_map(corp_test, content_transformer(removeNumbers))
```

```
## Warning in tm_map.SimpleCorpus(corp_test, content_transformer(removeNumbers)):
## transformation drops documents
```

```r
corp_test = tm_map(corp_test, content_transformer(removePunctuation))
```

```
## Warning in tm_map.SimpleCorpus(corp_test,
## content_transformer(removePunctuation)): transformation drops documents
```

```r
corp_test = tm_map(corp_test, content_transformer(stripWhitespace))
```

```
## Warning in tm_map.SimpleCorpus(corp_test, content_transformer(stripWhitespace)):
## transformation drops documents
```

```r
corp_test = tm_map(corp_test, content_transformer(removeWords),stopwords("en"))
```

```
## Warning in tm_map.SimpleCorpus(corp_test, content_transformer(removeWords), :
## transformation drops documents
```

```r
DTM_tst=DocumentTermMatrix(corp_test,list(dictionary=colnames(DTM_tr)))
tf_idf_tests = weightTfIdf(DTM_tst)
```

```
## Warning in weightTfIdf(DTM_tst): unreferenced term(s):
## modellingctrainaaronpressmannewsmltxt modellingctrainalancrosbynewsmltxt
```

```
## modellingctrainalexandersmithnewsmltxt modellingctrainbenjaminkanglimnewsmltxt
## modellingctrainbernardhickeynewsmltxt modellingctrainbraddorfmannewsmltxt
## modellingctraindarrenschuettlernewsmltxt modellingctraindavidlawdernewsmltxt
## modellingctrainednafernandesnewsmltxt modellingctrainericauchardnewsmltxt
## modellingctrainfumikofujisakinewsmltxt modellingctraingrahamearnshawnewsmltxt
## modellingctrainheatherscoffieldnewsmltxt modellingctrainjanlopatkanewsmltxt
## modellingctrainjanemacartneynewsmltxt modellingctrainjimgilchristnewsmltxt
## modellingctrainjowinterbottomnewsmltxt modellingctrainjoeortiznewsmltxt
## modellingctrainjohnmastrininewsmltxt modellingctrainjonathanbirtnewsmltxt
## modellingctrainkarlpenhaulnewsmltxt modellingctrainkeithweirnewsmltxt
## modellingctrainkevindrawbaughnewsmltxt modellingctrainkevinmorrisonnewsmltxt
## modellingctrainkirstinridleynewsmltxt modellingctrainkouroshkarimkhanynewsmltxt
## modellingctrainlydiazajcnewsmltxt modellingctrainlynneodonnellnewsmltxt
## modellingctrainlynnleybrowningnewsmltxt modellingctrainmarcelmichelsonnewsmltxt
## modellingctrainmarkbendeichnewsmltxt modellingctrainmartinwolknewsmltxt
## modellingctrainmatthewbuncenewsmltxt modellingctrainmichaelconnornewsmltxt
## modellingctrainmuredickienewsmltxt modellingctrainnicklouthnewsmltxt
## modellingctrainpatriciacomminsnewsmltxt modellingctrainpeterhumphreynewsmltxt
## modellingctrainpierretrannewsmltxt modellingctrainrobinsidelnewsmltxt
## modellingctrainrogerfillionnewsmltxt modellingctrainsamuelperrynewsmltxt
## modellingctrainsarahdavisonnewsmltxt modellingctrainscotthillisnewsmltxt
## modellingctrainsimoncowellnewsmltxt modellingctraintaneelynnewsmltxt
## modellingctraintheresepolettinewsmltxt modellingctraintimfarrandnewsmltxt
## modellingctraintoddnissennewsmltxt modellingctrainwilliamkazernewsmltxt
```

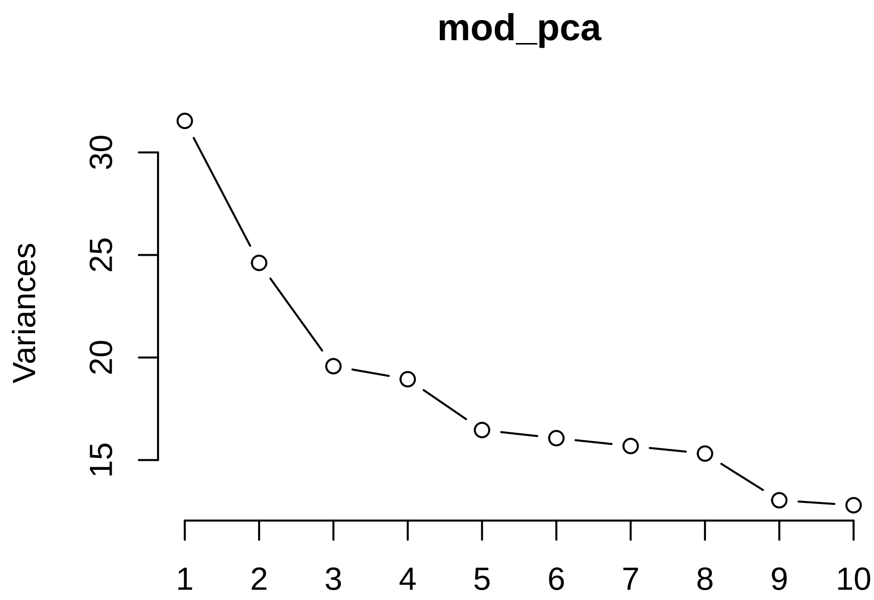
```r
DTM_test<-as.matrix(tf_idf_tests)
tf_idf_tests
```

```
## <<DocumentTermMatrix (documents: 2500, terms: 3395)>>
## Non-/sparse entries: 379314/8108186
## Sparsity           : 96%
## Maximal term length: 41
## Weighting          : term frequency - inverse document frequency (normalized) (tf-idf)
```

Now that we have created our train and test data sets, we need to reduce the dimensionality of our data so that we can apply a PCA analysis. This is related to the "curse of dimensionality" which is a problem because when our data because high dimensional, the models tend to start overfitting the data which leads to high out of sample error.

We then run a PCA using prcomp, scale=True to ensure that our input data has been standardized and has a zero mean and variance of one. We then plot our PCA outcome which shows a reduction in variance. We also need to find how much of the variance can be explained by each component. If we print out the proportion of variance to sum of variances, we see that at PC730 we have about 75% of variance explained by the components.

```r
DTM_train1<-DTM_train[,which(colSums(DTM_train) != 0)]
DTM_test1<-DTM_test[,which(colSums(DTM_test) != 0)]

DTM_test1 = DTM_test1[,intersect(colnames(DTM_test1),colnames(DTM_train1))]
DTM_train1 = DTM_train1[,intersect(colnames(DTM_test1),colnames(DTM_train1))]

mod_pca = prcomp(DTM_train1,scale=TRUE)
pred_pca=predict(mod_pca,newdata = DTM_test1)


plot(mod_pca,type='line')
```




```r
var <- apply(mod_pca$x, 2, var)
prop <- var / sum(var)
cumsum(prop)
```

```
##         PC1         PC2         PC3         PC4         PC5         PC6
## 0.009479095 0.016877628 0.022761820 0.028455367 0.033404000 0.038233810
##         PC7         PC8         PC9        PC10        PC11        PC12
## 0.042948219 0.047552109 0.051471990 0.055318969 0.058902425 0.062338432
##        PC13        PC14        PC15        PC16        PC17        PC18
## 0.065673432 0.068957926 0.072033770 0.075067784 0.078061800 0.080917395
##        PC19        PC20        PC21        PC22        PC23        PC24
## 0.083731736 0.086474435 0.089126133 0.091743850 0.094351409 0.096931852
##        PC25        PC26        PC27        PC28        PC29        PC30
## 0.099491097 0.101967086 0.104414891 0.106838750 0.109190550 0.111501594
##        PC31        PC32        PC33        PC34        PC35        PC36
## 0.113793732 0.116045716 0.118265526 0.120466087 0.122634062 0.124783760
##        PC37        PC38        PC39        PC40        PC41        PC42
## 0.126910899 0.129009644 0.131076087 0.133122077 0.135158511 0.137165685
##        PC43        PC44        PC45        PC46        PC47        PC48
## 0.139163086 0.141151843 0.143121313 0.145071606 0.146976444 0.148870374
##        PC49        PC50        PC51        PC52        PC53        PC54
## 0.150755299 0.152624633 0.154484678 0.156334196 0.158178721 0.160004063
##        PC55        PC56        PC57        PC58        PC59        PC60
## 0.161818326 0.163627203 0.165409070 0.167186673 0.168945938 0.170696810
##        PC61        PC62        PC63        PC64        PC65        PC66
## 0.172442699 0.174176534 0.175898114 0.177618009 0.179323692 0.181017078
##        PC67        PC68        PC69        PC70        PC71        PC72
## 0.182695354 0.184363673 0.186024337 0.187683897 0.189333429 0.190981127
##        PC73        PC74        PC75        PC76        PC77        PC78
```

```
## 0.192617224 0.194245216 0.195864624 0.197480624 0.199084826 0.200681120
##        PC79        PC80        PC81        PC82        PC83        PC84
## 0.202271907 0.203856480 0.205434195 0.207008458 0.208579343 0.210137522
##        PC85        PC86        PC87        PC88        PC89        PC90
## 0.211689618 0.213234863 0.214773338 0.216305930 0.217832990 0.219351109
##        PC91        PC92        PC93        PC94        PC95        PC96
## 0.220868076 0.222381416 0.223886219 0.225379815 0.226873030 0.228357379
##        PC97        PC98        PC99       PC100       PC101       PC102
## 0.229836167 0.231310251 0.232778140 0.234242558 0.235704150 0.237161285
##       PC103       PC104       PC105       PC106       PC107       PC108
## 0.238609851 0.240056907 0.241495977 0.242932536 0.244367392 0.245797406
##       PC109       PC110       PC111       PC112       PC113       PC114
## 0.247221644 0.248641440 0.250057820 0.251467213 0.252871416 0.254273537
##       PC115       PC116       PC117       PC118       PC119       PC120
## 0.255665537 0.257055487 0.258439869 0.259823038 0.261203216 0.262576989
##       PC121       PC122       PC123       PC124       PC125       PC126
## 0.263949648 0.265317615 0.266679961 0.268038133 0.269391758 0.270744190
##       PC127       PC128       PC129       PC130       PC131       PC132
## 0.272095199 0.273438503 0.274779883 0.276118807 0.277454749 0.278783499
##       PC133       PC134       PC135       PC136       PC137       PC138
## 0.280110914 0.281431815 0.282750472 0.284068034 0.285379301 0.286688321
##       PC139       PC140       PC141       PC142       PC143       PC144
## 0.287991801 0.289291227 0.290589417 0.291882250 0.293173735 0.294463775
##       PC145       PC146       PC147       PC148       PC149       PC150
## 0.295747933 0.297027319 0.298302619 0.299577067 0.300849159 0.302117852
##       PC151       PC152       PC153       PC154       PC155       PC156
## 0.303382876 0.304644887 0.305903898 0.307159565 0.308413367 0.309662569
##       PC157       PC158       PC159       PC160       PC161       PC162
## 0.310908236 0.312150264 0.313390155 0.314626279 0.315858274 0.317089463
##       PC163       PC164       PC165       PC166       PC167       PC168
## 0.318315336 0.319539551 0.320760321 0.321977906 0.323191907 0.324403621
##       PC169       PC170       PC171       PC172       PC173       PC174
## 0.325614165 0.326818656 0.328021552 0.329220602 0.330418991 0.331613609
##       PC175       PC176       PC177       PC178       PC179       PC180
## 0.332805167 0.333993229 0.335179933 0.336361186 0.337541040 0.338717731
##       PC181       PC182       PC183       PC184       PC185       PC186
## 0.339892047 0.341065532 0.342234619 0.343402911 0.344568857 0.345732929
##       PC187       PC188       PC189       PC190       PC191       PC192
## 0.346891653 0.348046059 0.349198767 0.350349477 0.351498578 0.352643619
##       PC193       PC194       PC195       PC196       PC197       PC198
## 0.353787875 0.354930412 0.356071967 0.357209324 0.358344185 0.359477832
##       PC199       PC200       PC201       PC202       PC203       PC204
## 0.360605970 0.361732731 0.362859378 0.363983727 0.365103589 0.366220848
##       PC205       PC206       PC207       PC208       PC209       PC210
## 0.367333572 0.368445090 0.369555752 0.370664165 0.371771478 0.372874896
##       PC211       PC212       PC213       PC214       PC215       PC216
## 0.373976968 0.375077967 0.376176299 0.377271439 0.378364798 0.379457219
##       PC217       PC218       PC219       PC220       PC221       PC222
## 0.380545196 0.381632276 0.382715524 0.383796961 0.384875186 0.385951129
##       PC223       PC224       PC225       PC226       PC227       PC228
## 0.387025370 0.388097351 0.389168342 0.390237362 0.391305673 0.392372282
##       PC229       PC230       PC231       PC232       PC233       PC234
## 0.393435807 0.394497449 0.395558375 0.396615904 0.397670913 0.398724214
##       PC235       PC236       PC237       PC238       PC239       PC240
```

```
## 0.399774591 0.400820057 0.401864988 0.402908814 0.403950546 0.404989074
##       PC241       PC242       PC243       PC244       PC245       PC246
## 0.406025205 0.407059856 0.408092258 0.409122788 0.410149609 0.411175269
##       PC247       PC248       PC249       PC250       PC251       PC252
## 0.412198675 0.413218808 0.414237722 0.415255474 0.416271794 0.417287057
##       PC253       PC254       PC255       PC256       PC257       PC258
## 0.418300955 0.419313087 0.420323326 0.421330178 0.422335418 0.423338929
##       PC259       PC260       PC261       PC262       PC263       PC264
## 0.424339320 0.425337366 0.426334166 0.427328665 0.428321215 0.429312721
##       PC265       PC266       PC267       PC268       PC269       PC270
## 0.430301963 0.431290637 0.432277505 0.433261741 0.434243074 0.435223243
##       PC271       PC272       PC273       PC274       PC275       PC276
## 0.436202043 0.437179145 0.438154237 0.439127506 0.440098518 0.441068704
##       PC277       PC278       PC279       PC280       PC281       PC282
## 0.442035513 0.442998544 0.443960956 0.444922037 0.445881958 0.446841178
##       PC283       PC284       PC285       PC286       PC287       PC288
## 0.447798773 0.448754931 0.449708359 0.450660294 0.451611449 0.452559781
##       PC289       PC290       PC291       PC292       PC293       PC294
## 0.453507335 0.454454369 0.455399715 0.456342416 0.457284213 0.458223378
##       PC295       PC296       PC297       PC298       PC299       PC300
## 0.459160463 0.460094515 0.461026156 0.461957492 0.462886723 0.463814576
##       PC301       PC302       PC303       PC304       PC305       PC306
## 0.464741493 0.465667309 0.466591273 0.467514202 0.468435191 0.469353247
##       PC307       PC308       PC309       PC310       PC311       PC312
## 0.470270475 0.471185515 0.472098631 0.473010935 0.473920012 0.474828119
##       PC313       PC314       PC315       PC316       PC317       PC318
## 0.475735509 0.476641815 0.477547766 0.478451000 0.479352256 0.480251571
##       PC319       PC320       PC321       PC322       PC323       PC324
## 0.481150490 0.482047184 0.482943568 0.483838671 0.484732430 0.485623104
##       PC325       PC326       PC327       PC328       PC329       PC330
## 0.486512576 0.487400095 0.488287177 0.489173330 0.490057113 0.490938905
##       PC331       PC332       PC333       PC334       PC335       PC336
## 0.491819773 0.492698882 0.493577054 0.494453077 0.495327797 0.496199038
##       PC337       PC338       PC339       PC340       PC341       PC342
## 0.497069330 0.497938386 0.498806342 0.499673219 0.500539513 0.501404464
##       PC343       PC344       PC345       PC346       PC347       PC348
## 0.502267345 0.503128318 0.503988475 0.504845375 0.505701223 0.506556958
##       PC349       PC350       PC351       PC352       PC353       PC354
## 0.507410776 0.508262021 0.509112512 0.509961168 0.510808092 0.511653500
##       PC355       PC356       PC357       PC358       PC359       PC360
## 0.512498113 0.513341190 0.514183048 0.515022316 0.515861371 0.516696779
##       PC361       PC362       PC363       PC364       PC365       PC366
## 0.517531625 0.518364571 0.519196027 0.520027103 0.520857123 0.521685932
##       PC367       PC368       PC369       PC370       PC371       PC372
## 0.522512310 0.523338583 0.524162920 0.524985169 0.525804513 0.526623591
##       PC373       PC374       PC375       PC376       PC377       PC378
## 0.527440740 0.528257051 0.529072064 0.529885732 0.530699117 0.531511474
##       PC379       PC380       PC381       PC382       PC383       PC384
## 0.532322266 0.533132508 0.533939504 0.534744984 0.535549683 0.536352366
##       PC385       PC386       PC387       PC388       PC389       PC390
## 0.537154100 0.537955214 0.538754618 0.539553114 0.540350919 0.541147461
##       PC391       PC392       PC393       PC394       PC395       PC396
## 0.541940819 0.542733143 0.543524960 0.544314586 0.545102770 0.545889993
##       PC397       PC398       PC399       PC400       PC401       PC402
```

```
## 0.546676524 0.547461482 0.548246114 0.549029292 0.549810830 0.550591289
##       PC403       PC404       PC405       PC406       PC407       PC408
## 0.551369881 0.552147774 0.552923045 0.553697734 0.554472056 0.555245803
##       PC409       PC410       PC411       PC412       PC413       PC414
## 0.556018232 0.556788434 0.557557834 0.558325245 0.559092079 0.559858507
##       PC415       PC416       PC417       PC418       PC419       PC420
## 0.560623611 0.561388274 0.562151240 0.562913484 0.563674135 0.564433261
##       PC421       PC422       PC423       PC424       PC425       PC426
## 0.565190778 0.565947052 0.566703004 0.567455391 0.568206962 0.568957749
##       PC427       PC428       PC429       PC430       PC431       PC432
## 0.569706608 0.570454533 0.571201755 0.571946876 0.572691310 0.573435250
##       PC433       PC434       PC435       PC436       PC437       PC438
## 0.574177297 0.574918610 0.575659030 0.576398029 0.577136093 0.577873343
##       PC439       PC440       PC441       PC442       PC443       PC444
## 0.578608913 0.579344207 0.580078084 0.580810443 0.581541879 0.582272789
##       PC445       PC446       PC447       PC448       PC449       PC450
## 0.583002659 0.583730458 0.584457762 0.585184298 0.585909113 0.586632010
##       PC451       PC452       PC453       PC454       PC455       PC456
## 0.587354610 0.588075998 0.588795819 0.589515227 0.590233555 0.590950867
##       PC457       PC458       PC459       PC460       PC461       PC462
## 0.591666793 0.592380216 0.593092717 0.593804894 0.594515339 0.595224481
##       PC463       PC464       PC465       PC466       PC467       PC468
## 0.595931593 0.596638143 0.597343899 0.598048038 0.598751362 0.599453790
##       PC469       PC470       PC471       PC472       PC473       PC474
## 0.600154696 0.600854806 0.601554344 0.602251286 0.602947099 0.603642078
##       PC475       PC476       PC477       PC478       PC479       PC480
## 0.604336943 0.605030228 0.605723173 0.606414055 0.607104418 0.607793715
##       PC481       PC482       PC483       PC484       PC485       PC486
## 0.608481418 0.609168688 0.609854702 0.610540365 0.611224295 0.611906239
##       PC487       PC488       PC489       PC490       PC491       PC492
## 0.612587229 0.613267711 0.613945699 0.614622724 0.615299406 0.615974893
##       PC493       PC494       PC495       PC496       PC497       PC498
## 0.616649843 0.617322729 0.617994763 0.618666467 0.619338022 0.620006738
##       PC499       PC500       PC501       PC502       PC503       PC504
## 0.620675045 0.621342134 0.622008429 0.622672804 0.623336064 0.623999026
##       PC505       PC506       PC507       PC508       PC509       PC510
## 0.624661399 0.625322613 0.625982796 0.626640864 0.627298479 0.627955001
##       PC511       PC512       PC513       PC514       PC515       PC516
## 0.628611480 0.629265408 0.629918502 0.630571148 0.631222857 0.631873389
##       PC517       PC518       PC519       PC520       PC521       PC522
## 0.632522768 0.633171913 0.633819781 0.634466982 0.635113392 0.635757959
##       PC523       PC524       PC525       PC526       PC527       PC528
## 0.636401465 0.637044624 0.637686851 0.638327934 0.638968505 0.639607262
##       PC529       PC530       PC531       PC532       PC533       PC534
## 0.640245260 0.640881621 0.641516662 0.642150937 0.642783584 0.643415763
##       PC535       PC536       PC537       PC538       PC539       PC540
## 0.644047632 0.644678422 0.645309061 0.645938387 0.646566799 0.647194799
##       PC541       PC542       PC543       PC544       PC545       PC546
## 0.647820985 0.648445572 0.649069478 0.649692045 0.650314463 0.650936210
##       PC547       PC548       PC549       PC550       PC551       PC552
## 0.651556329 0.652175820 0.652794207 0.653412350 0.654028338 0.654643904
##       PC553       PC554       PC555       PC556       PC557       PC558
## 0.655258785 0.655872826 0.656485324 0.657096950 0.657707794 0.658318295
##       PC559       PC560       PC561       PC562       PC563       PC564
```

```
## 0.658927637 0.659536172 0.660144597 0.660752080 0.661358549 0.661963891
##       PC565       PC566       PC567       PC568       PC569       PC570
## 0.662567522 0.663170367 0.663772787 0.664373906 0.664974130 0.665572956
##       PC571       PC572       PC573       PC574       PC575       PC576
## 0.666171462 0.666768400 0.667364955 0.667960376 0.668555366 0.669149733
##       PC577       PC578       PC579       PC580       PC581       PC582
## 0.669742348 0.670334721 0.670925299 0.671515198 0.672104217 0.672692594
##       PC583       PC584       PC585       PC586       PC587       PC588
## 0.673279946 0.673866705 0.674452799 0.675037084 0.675621027 0.676204269
##       PC589       PC590       PC591       PC592       PC593       PC594
## 0.676785714 0.677366149 0.677944756 0.678522723 0.679100358 0.679677066
##       PC595       PC596       PC597       PC598       PC599       PC600
## 0.680253177 0.680828735 0.681403354 0.681976948 0.682550405 0.683122656
##       PC601       PC602       PC603       PC604       PC605       PC606
## 0.683693926 0.684264557 0.684834936 0.685404716 0.685972401 0.686539840
##       PC607       PC608       PC609       PC610       PC611       PC612
## 0.687106833 0.687672783 0.688237186 0.688800431 0.689363235 0.689924969
##       PC613       PC614       PC615       PC616       PC617       PC618
## 0.690485841 0.691045446 0.691604413 0.692162297 0.692719421 0.693275264
##       PC619       PC620       PC621       PC622       PC623       PC624
## 0.693830812 0.694386245 0.694940017 0.695492811 0.696044835 0.696596132
##       PC625       PC626       PC627       PC628       PC629       PC630
## 0.697147092 0.697696743 0.698245397 0.698793321 0.699340902 0.699888007
##       PC631       PC632       PC633       PC634       PC635       PC636
## 0.700434342 0.700979872 0.701525226 0.702069846 0.702613296 0.703155607
##       PC637       PC638       PC639       PC640       PC641       PC642
## 0.703696583 0.704236364 0.704775599 0.705313359 0.705850719 0.706387235
##       PC643       PC644       PC645       PC646       PC647       PC648
## 0.706923388 0.707458813 0.707993801 0.708527922 0.709061215 0.709594187
##       PC649       PC650       PC651       PC652       PC653       PC654
## 0.710125772 0.710656485 0.711186307 0.711715755 0.712243278 0.712770048
##       PC655       PC656       PC657       PC658       PC659       PC660
## 0.713296260 0.713822274 0.714347424 0.714871831 0.715395055 0.715917134
##       PC661       PC662       PC663       PC664       PC665       PC666
## 0.716438701 0.716958674 0.717478429 0.717997882 0.718516264 0.719033948
##       PC667       PC668       PC669       PC670       PC671       PC672
## 0.719550714 0.720066457 0.720581186 0.721095209 0.721608420 0.722121084
##       PC673       PC674       PC675       PC676       PC677       PC678
## 0.722633166 0.723143283 0.723652880 0.724161710 0.724670367 0.725177681
##       PC679       PC680       PC681       PC682       PC683       PC684
## 0.725684197 0.726190326 0.726696170 0.727201762 0.727706435 0.728210097
##       PC685       PC686       PC687       PC688       PC689       PC690
## 0.728713115 0.729215416 0.729716266 0.730216766 0.730716562 0.731215329
##       PC691       PC692       PC693       PC694       PC695       PC696
## 0.731713594 0.732211398 0.732708636 0.733205313 0.733700484 0.734195354
##       PC697       PC698       PC699       PC700       PC701       PC702
## 0.734689205 0.735181978 0.735674155 0.736165665 0.736656713 0.737146339
##       PC703       PC704       PC705       PC706       PC707       PC708
## 0.737635834 0.738124214 0.738612192 0.739099376 0.739586086 0.740072453
##       PC709       PC710       PC711       PC712       PC713       PC714
## 0.740557458 0.741041570 0.741525072 0.742008158 0.742490828 0.742972148
##       PC715       PC716       PC717       PC718       PC719       PC720
## 0.743452849 0.743932891 0.744412058 0.744890577 0.745368040 0.745844561
##       PC721       PC722       PC723       PC724       PC725       PC726
```

```
## 0.746320232 0.746795429 0.747269815 0.747743559 0.748216572 0.748689010
##       PC727       PC728       PC729       PC730       PC731       PC732
## 0.749160520 0.749631692 0.750102194 0.750572353 0.751041441 0.751509553
##       PC733       PC734       PC735       PC736       PC737       PC738
## 0.751976768 0.752443465 0.752909758 0.753374366 0.753838416 0.754302068
##       PC739       PC740       PC741       PC742       PC743       PC744
## 0.754764447 0.755226515 0.755688102 0.756149474 0.756609702 0.757069313
##       PC745       PC746       PC747       PC748       PC749       PC750
## 0.757528468 0.757987217 0.758444967 0.758902352 0.759359348 0.759815862
##       PC751       PC752       PC753       PC754       PC755       PC756
## 0.760271347 0.760725635 0.761179434 0.761632929 0.762085818 0.762537292
##       PC757       PC758       PC759       PC760       PC761       PC762
## 0.762988298 0.763439189 0.763889378 0.764339188 0.764788664 0.765236858
##       PC763       PC764       PC765       PC766       PC767       PC768
## 0.765684641 0.766131528 0.766576833 0.767021911 0.767465982 0.767909755
##       PC769       PC770       PC771       PC772       PC773       PC774
## 0.768353010 0.768795616 0.769237500 0.769678311 0.770118940 0.770558361
##       PC775       PC776       PC777       PC778       PC779       PC780
## 0.770996799 0.771434891 0.771872376 0.772309707 0.772746293 0.773182117
##       PC781       PC782       PC783       PC784       PC785       PC786
## 0.773617085 0.774051688 0.774485507 0.774918885 0.775351124 0.775782472
##       PC787       PC788       PC789       PC790       PC791       PC792
## 0.776213075 0.776643296 0.777073434 0.777502620 0.777931272 0.778359293
##       PC793       PC794       PC795       PC796       PC797       PC798
## 0.778786456 0.779212588 0.779638149 0.780062263 0.780486261 0.780909876
##       PC799       PC800       PC801       PC802       PC803       PC804
## 0.781332965 0.781755061 0.782176664 0.782597343 0.783016772 0.783435790
##       PC805       PC806       PC807       PC808       PC809       PC810
## 0.783854039 0.784271379 0.784688630 0.785105244 0.785521453 0.785937159
##       PC811       PC812       PC813       PC814       PC815       PC816
## 0.786351771 0.786766068 0.787179762 0.787593282 0.788006232 0.788418621
##       PC817       PC818       PC819       PC820       PC821       PC822
## 0.788830129 0.789241013 0.789651362 0.790061223 0.790470385 0.790878911
##       PC823       PC824       PC825       PC826       PC827       PC828
## 0.791286189 0.791693053 0.792099458 0.792505487 0.792910937 0.793314846
##       PC829       PC830       PC831       PC832       PC833       PC834
## 0.793718237 0.794121209 0.794524005 0.794926137 0.795327257 0.795727657
##       PC835       PC836       PC837       PC838       PC839       PC840
## 0.796127735 0.796526944 0.796925818 0.797324360 0.797722016 0.798118597
##       PC841       PC842       PC843       PC844       PC845       PC846
## 0.798515075 0.798911370 0.799307173 0.799701503 0.800095616 0.800488805
##       PC847       PC848       PC849       PC850       PC851       PC852
## 0.800881923 0.801274604 0.801666680 0.802057972 0.802448320 0.802838497
##       PC853       PC854       PC855       PC856       PC857       PC858
## 0.803228111 0.803617271 0.804006046 0.804394556 0.804782185 0.805168845
##       PC859       PC860       PC861       PC862       PC863       PC864
## 0.805555174 0.805940493 0.806325272 0.806709237 0.807093003 0.807475888
##       PC865       PC866       PC867       PC868       PC869       PC870
## 0.807858555 0.808240116 0.808621598 0.809002644 0.809382528 0.809762036
##       PC871       PC872       PC873       PC874       PC875       PC876
## 0.810140514 0.810518498 0.810895905 0.811272971 0.811649410 0.812025697
##       PC877       PC878       PC879       PC880       PC881       PC882
## 0.812401128 0.812775828 0.813150319 0.813524312 0.813897771 0.814270162
##       PC883       PC884       PC885       PC886       PC887       PC888
```

```
## 0.814642117 0.815013576 0.815384315 0.815754747 0.816124386 0.816493442
##       PC889       PC890       PC891       PC892       PC893       PC894
## 0.816861526 0.817228973 0.817596210 0.817962865 0.818328774 0.818693786
##       PC895       PC896       PC897       PC898       PC899       PC900
## 0.819058345 0.819422436 0.819786109 0.820149553 0.820511979 0.820873859
##       PC901       PC902       PC903       PC904       PC905       PC906
## 0.821234789 0.821594956 0.821954860 0.822314476 0.822673530 0.823031935
##       PC907       PC908       PC909       PC910       PC911       PC912
## 0.823389767 0.823747218 0.824104046 0.824460029 0.824815632 0.825170742
##       PC913       PC914       PC915       PC916       PC917       PC918
## 0.825525381 0.825879688 0.826233255 0.826586655 0.826939398 0.827291763
##       PC919       PC920       PC921       PC922       PC923       PC924
## 0.827643466 0.827994744 0.828345221 0.828695340 0.829044407 0.829393310
##       PC925       PC926       PC927       PC928       PC929       PC930
## 0.829741725 0.830090123 0.830437736 0.830784731 0.831131359 0.831476879
##       PC931       PC932       PC933       PC934       PC935       PC936
## 0.831822372 0.832166804 0.832510911 0.832854803 0.833197824 0.833539449
##       PC937       PC938       PC939       PC940       PC941       PC942
## 0.833880939 0.834221669 0.834562192 0.834902204 0.835241686 0.835580519
##       PC943       PC944       PC945       PC946       PC947       PC948
## 0.835918468 0.836256178 0.836593357 0.836930286 0.837266579 0.837602186
##       PC949       PC950       PC951       PC952       PC953       PC954
## 0.837936720 0.838270986 0.838604793 0.838938508 0.839271259 0.839603475
##       PC955       PC956       PC957       PC958       PC959       PC960
## 0.839935062 0.840266230 0.840596800 0.840927022 0.841257006 0.841586040
##       PC961       PC962       PC963       PC964       PC965       PC966
## 0.841914817 0.842243105 0.842571182 0.842898447 0.843224748 0.843550106
##       PC967       PC968       PC969       PC970       PC971       PC972
## 0.843875152 0.844200010 0.844524245 0.844847871 0.845171196 0.845494390
##       PC973       PC974       PC975       PC976       PC977       PC978
## 0.845816807 0.846138710 0.846459422 0.846779811 0.847099910 0.847419758
##       PC979       PC980       PC981       PC982       PC983       PC984
## 0.847738620 0.848057418 0.848375948 0.848694077 0.849010882 0.849327657
##       PC985       PC986       PC987       PC988       PC989       PC990
## 0.849644083 0.849960066 0.850275262 0.850589752 0.850903779 0.851217634
##       PC991       PC992       PC993       PC994       PC995       PC996
## 0.851531211 0.851844372 0.852157103 0.852469210 0.852780945 0.853092231
##       PC997       PC998       PC999      PC1000      PC1001      PC1002
## 0.853402836 0.853713011 0.854022370 0.854331414 0.854639790 0.854947938
##      PC1003      PC1004      PC1005      PC1006      PC1007      PC1008
## 0.855255506 0.855562889 0.855869585 0.856176151 0.856482469 0.856788172
##      PC1009      PC1010      PC1011      PC1012      PC1013      PC1014
## 0.857093362 0.857398365 0.857702973 0.858006886 0.858310244 0.858612972
##      PC1015      PC1016      PC1017      PC1018      PC1019      PC1020
## 0.858914806 0.859216332 0.859517509 0.859818435 0.860119008 0.860418556
##      PC1021      PC1022      PC1023      PC1024      PC1025      PC1026
## 0.860716982 0.861015309 0.861313115 0.861610600 0.861907699 0.862204632
##      PC1027      PC1028      PC1029      PC1030      PC1031      PC1032
## 0.862500699 0.862796476 0.863091740 0.863386517 0.863681122 0.863975194
##      PC1033      PC1034      PC1035      PC1036      PC1037      PC1038
## 0.864268826 0.864561924 0.864854555 0.865146377 0.865437861 0.865728906
##      PC1039      PC1040      PC1041      PC1042      PC1043      PC1044
## 0.866019572 0.866309652 0.866599467 0.866888534 0.867177099 0.867465498
##      PC1045      PC1046      PC1047      PC1048      PC1049      PC1050
```

```
## 0.867753310 0.868040284 0.868326777 0.868612878 0.868898640 0.869184226
##      PC1051      PC1052      PC1053      PC1054      PC1055      PC1056
## 0.869469434 0.869753862 0.870037664 0.870321144 0.870604389 0.870887052
##      PC1057      PC1058      PC1059      PC1060      PC1061      PC1062
## 0.871169153 0.871450448 0.871731293 0.872011726 0.872291752 0.872571335
##      PC1063      PC1064      PC1065      PC1066      PC1067      PC1068
## 0.872850734 0.873129742 0.873408217 0.873686055 0.873963526 0.874240555
##      PC1069      PC1070      PC1071      PC1072      PC1073      PC1074
## 0.874516993 0.874793042 0.875068634 0.875343814 0.875618454 0.875892443
##      PC1075      PC1076      PC1077      PC1078      PC1079      PC1080
## 0.876166384 0.876439563 0.876712355 0.876985016 0.877257259 0.877529067
##      PC1081      PC1082      PC1083      PC1084      PC1085      PC1086
## 0.877800564 0.878071676 0.878341984 0.878611930 0.878881542 0.879150810
##      PC1087      PC1088      PC1089      PC1090      PC1091      PC1092
## 0.879419774 0.879687853 0.879955654 0.880223184 0.880489583 0.880755688
##      PC1093      PC1094      PC1095      PC1096      PC1097      PC1098
## 0.881021540 0.881286896 0.881551454 0.881815878 0.882079895 0.882343586
##      PC1099      PC1100      PC1101      PC1102      PC1103      PC1104
## 0.882607023 0.882870039 0.883132196 0.883394178 0.883655132 0.883915817
##      PC1105      PC1106      PC1107      PC1108      PC1109      PC1110
## 0.884176293 0.884436475 0.884695866 0.884955165 0.885214120 0.885472532
##      PC1111      PC1112      PC1113      PC1114      PC1115      PC1116
## 0.885730698 0.885988448 0.886245772 0.886502392 0.886758423 0.887013959
##      PC1117      PC1118      PC1119      PC1120      PC1121      PC1122
## 0.887269341 0.887524217 0.887778715 0.888033038 0.888287154 0.888540725
##      PC1123      PC1124      PC1125      PC1126      PC1127      PC1128
## 0.888793957 0.889046590 0.889299095 0.889550864 0.889802589 0.890053795
##      PC1129      PC1130      PC1131      PC1132      PC1133      PC1134
## 0.890304537 0.890554683 0.890804750 0.891054527 0.891303638 0.891552161
##      PC1135      PC1136      PC1137      PC1138      PC1139      PC1140
## 0.891800317 0.892048214 0.892295845 0.892542924 0.892789669 0.893035740
##      PC1141      PC1142      PC1143      PC1144      PC1145      PC1146
## 0.893281399 0.893526812 0.893772010 0.894016409 0.894260556 0.894504404
##      PC1147      PC1148      PC1149      PC1150      PC1151      PC1152
## 0.894747525 0.894990393 0.895232654 0.895474509 0.895716188 0.895957349
##      PC1153      PC1154      PC1155      PC1156      PC1157      PC1158
## 0.896198396 0.896438969 0.896678508 0.896917741 0.897156898 0.897395517
##      PC1159      PC1160      PC1161      PC1162      PC1163      PC1164
## 0.897633561 0.897871354 0.898108865 0.898346126 0.898582580 0.898818573
##      PC1165      PC1166      PC1167      PC1168      PC1169      PC1170
## 0.899054468 0.899290176 0.899525541 0.899760582 0.899994941 0.900229067
##      PC1171      PC1172      PC1173      PC1174      PC1175      PC1176
## 0.900462639 0.900695434 0.900928077 0.901160516 0.901392758 0.901624418
##      PC1177      PC1178      PC1179      PC1180      PC1181      PC1182
## 0.901855611 0.902086748 0.902317626 0.902547698 0.902777579 0.903007321
##      PC1183      PC1184      PC1185      PC1186      PC1187      PC1188
## 0.903236799 0.903465933 0.903694536 0.903922806 0.904150714 0.904377931
##      PC1189      PC1190      PC1191      PC1192      PC1193      PC1194
## 0.904604843 0.904831202 0.905057076 0.905282564 0.905507528 0.905732062
##      PC1195      PC1196      PC1197      PC1198      PC1199      PC1200
## 0.905955945 0.906179207 0.906402104 0.906624718 0.906847187 0.907069229
##      PC1201      PC1202      PC1203      PC1204      PC1205      PC1206
## 0.907290810 0.907512340 0.907732912 0.907953355 0.908173347 0.908393109
##      PC1207      PC1208      PC1209      PC1210      PC1211      PC1212
```

```
## 0.908612540 0.908831236 0.909049642 0.909267821 0.909485555 0.909703051
##      PC1213      PC1214      PC1215      PC1216      PC1217      PC1218
## 0.909919955 0.910136678 0.910352805 0.910568793 0.910784688 0.911000103
##      PC1219      PC1220      PC1221      PC1222      PC1223      PC1224
## 0.911214839 0.911429440 0.911643788 0.911857656 0.912071292 0.912284673
##      PC1225      PC1226      PC1227      PC1228      PC1229      PC1230
## 0.912497494 0.912710023 0.912922386 0.913133978 0.913344774 0.913555246
##      PC1231      PC1232      PC1233      PC1234      PC1235      PC1236
## 0.913765395 0.913975434 0.914185364 0.914394528 0.914603500 0.914812245
##      PC1237      PC1238      PC1239      PC1240      PC1241      PC1242
## 0.915020518 0.915228693 0.915436072 0.915643352 0.915850358 0.916056821
##      PC1243      PC1244      PC1245      PC1246      PC1247      PC1248
## 0.916262854 0.916468296 0.916673311 0.916878160 0.917082900 0.917286798
##      PC1249      PC1250      PC1251      PC1252      PC1253      PC1254
## 0.917490485 0.917693947 0.917897119 0.918100185 0.918302605 0.918504315
##      PC1255      PC1256      PC1257      PC1258      PC1259      PC1260
## 0.918705950 0.918907292 0.919108275 0.919308967 0.919509323 0.919709267
##      PC1261      PC1262      PC1263      PC1264      PC1265      PC1266
## 0.919908790 0.920107757 0.920306602 0.920505119 0.920703500 0.920901164
##      PC1267      PC1268      PC1269      PC1270      PC1271      PC1272
## 0.921098638 0.921295902 0.921492761 0.921689434 0.921885757 0.922081780
##      PC1273      PC1274      PC1275      PC1276      PC1277      PC1278
## 0.922277615 0.922473097 0.922668459 0.922863564 0.923057833 0.923251820
##      PC1279      PC1280      PC1281      PC1282      PC1283      PC1284
## 0.923445183 0.923638306 0.923831375 0.924024112 0.924216423 0.924408425
##      PC1285      PC1286      PC1287      PC1288      PC1289      PC1290
## 0.924600000 0.924791242 0.924981915 0.925172418 0.925362598 0.925552182
##      PC1291      PC1292      PC1293      PC1294      PC1295      PC1296
## 0.925741618 0.925930464 0.926119093 0.926307588 0.926495249 0.926682692
##      PC1297      PC1298      PC1299      PC1300      PC1301      PC1302
## 0.926870002 0.927057242 0.927243893 0.927430042 0.927616032 0.927801567
##      PC1303      PC1304      PC1305      PC1306      PC1307      PC1308
## 0.927986713 0.928171703 0.928356494 0.928541003 0.928724668 0.928908228
##      PC1309      PC1310      PC1311      PC1312      PC1313      PC1314
## 0.929091437 0.929274207 0.929456771 0.929638897 0.929820787 0.930002048
##      PC1315      PC1316      PC1317      PC1318      PC1319      PC1320
## 0.930182994 0.930363669 0.930543844 0.930723916 0.930903465 0.931082881
##      PC1321      PC1322      PC1323      PC1324      PC1325      PC1326
## 0.931262080 0.931441179 0.931619752 0.931797834 0.931975900 0.932153648
##      PC1327      PC1328      PC1329      PC1330      PC1331      PC1332
## 0.932330858 0.932507672 0.932684119 0.932860153 0.933036012 0.933211465
##      PC1333      PC1334      PC1335      PC1336      PC1337      PC1338
## 0.933386444 0.933561104 0.933735582 0.933909638 0.934083167 0.934256458
##      PC1339      PC1340      PC1341      PC1342      PC1343      PC1344
## 0.934429387 0.934602272 0.934774703 0.934947060 0.935118810 0.935290495
##      PC1345      PC1346      PC1347      PC1348      PC1349      PC1350
## 0.935461786 0.935633018 0.935803814 0.935974362 0.936144493 0.936314280
##      PC1351      PC1352      PC1353      PC1354      PC1355      PC1356
## 0.936483585 0.936652656 0.936821697 0.936990195 0.937158225 0.937325755
##      PC1357      PC1358      PC1359      PC1360      PC1361      PC1362
## 0.937492967 0.937659892 0.937826766 0.937993297 0.938159686 0.938325711
##      PC1363      PC1364      PC1365      PC1366      PC1367      PC1368
## 0.938491225 0.938656553 0.938821501 0.938986204 0.939150573 0.939314633
##      PC1369      PC1370      PC1371      PC1372      PC1373      PC1374
```

```
## 0.939478404 0.939641819 0.939805201 0.939967876 0.940130253 0.940292446
##      PC1375      PC1376      PC1377      PC1378      PC1379      PC1380
## 0.940454564 0.940615989 0.940776957 0.940937563 0.941097729 0.941257870
##      PC1381      PC1382      PC1383      PC1384      PC1385      PC1386
## 0.941417843 0.941577535 0.941736785 0.941895756 0.942054562 0.942213136
##      PC1387      PC1388      PC1389      PC1390      PC1391      PC1392
## 0.942371412 0.942529165 0.942686734 0.942843991 0.943000888 0.943157599
##      PC1393      PC1394      PC1395      PC1396      PC1397      PC1398
## 0.943314021 0.943470224 0.943625823 0.943781162 0.943935916 0.944090354
##      PC1399      PC1400      PC1401      PC1402      PC1403      PC1404
## 0.944244599 0.944398722 0.944552359 0.944705801 0.944858868 0.945011845
##      PC1405      PC1406      PC1407      PC1408      PC1409      PC1410
## 0.945164588 0.945317169 0.945469371 0.945621289 0.945772732 0.945923836
##      PC1411      PC1412      PC1413      PC1414      PC1415      PC1416
## 0.946074362 0.946224737 0.946374931 0.946524912 0.946674465 0.946823806
##      PC1417      PC1418      PC1419      PC1420      PC1421      PC1422
## 0.946972829 0.947121475 0.947269809 0.947417939 0.947565728 0.947713239
##      PC1423      PC1424      PC1425      PC1426      PC1427      PC1428
## 0.947860571 0.948007696 0.948154502 0.948300899 0.948447213 0.948593095
##      PC1429      PC1430      PC1431      PC1432      PC1433      PC1434
## 0.948738824 0.948884040 0.949029087 0.949174071 0.949318663 0.949462880
##      PC1435      PC1436      PC1437      PC1438      PC1439      PC1440
## 0.949607045 0.949750761 0.949894058 0.950037103 0.950180000 0.950322660
##      PC1441      PC1442      PC1443      PC1444      PC1445      PC1446
## 0.950465240 0.950607331 0.950749328 0.950891045 0.951032224 0.951173316
##      PC1447      PC1448      PC1449      PC1450      PC1451      PC1452
## 0.951313805 0.951454103 0.951594063 0.951733712 0.951873014 0.952012134
##      PC1453      PC1454      PC1455      PC1456      PC1457      PC1458
## 0.952150742 0.952288994 0.952427189 0.952564987 0.952702442 0.952839524
##      PC1459      PC1460      PC1461      PC1462      PC1463      PC1464
## 0.952976154 0.953112554 0.953248830 0.953384986 0.953520723 0.953656302
##      PC1465      PC1466      PC1467      PC1468      PC1469      PC1470
## 0.953791690 0.953926718 0.954061535 0.954196144 0.954330563 0.954464655
##      PC1471      PC1472      PC1473      PC1474      PC1475      PC1476
## 0.954598415 0.954732067 0.954865411 0.954998640 0.955131305 0.955263722
##      PC1477      PC1478      PC1479      PC1480      PC1481      PC1482
## 0.955395736 0.955527339 0.955658616 0.955789644 0.955920526 0.956051315
##      PC1483      PC1484      PC1485      PC1486      PC1487      PC1488
## 0.956181468 0.956311544 0.956441406 0.956570908 0.956700209 0.956829384
##      PC1489      PC1490      PC1491      PC1492      PC1493      PC1494
## 0.956958224 0.957086876 0.957215462 0.957343521 0.957471297 0.957598798
##      PC1495      PC1496      PC1497      PC1498      PC1499      PC1500
## 0.957726060 0.957853054 0.957979860 0.958106480 0.958232672 0.958358587
##      PC1501      PC1502      PC1503      PC1504      PC1505      PC1506
## 0.958484227 0.958609587 0.958734650 0.958859555 0.958984254 0.959108654
##      PC1507      PC1508      PC1509      PC1510      PC1511      PC1512
## 0.959232832 0.959356819 0.959480542 0.959604091 0.959727428 0.959850381
##      PC1513      PC1514      PC1515      PC1516      PC1517      PC1518
## 0.959973127 0.960095670 0.960217681 0.960339295 0.960460688 0.960581821
##      PC1519      PC1520      PC1521      PC1522      PC1523      PC1524
## 0.960702829 0.960823785 0.960944512 0.961065030 0.961184989 0.961304814
##      PC1525      PC1526      PC1527      PC1528      PC1529      PC1530
## 0.961424309 0.961543520 0.961662521 0.961781298 0.961899784 0.962018190
##      PC1531      PC1532      PC1533      PC1534      PC1535      PC1536
```

```
## 0.962136199 0.962253960 0.962371461 0.962488751 0.962605937 0.962722633
##      PC1537      PC1538      PC1539      PC1540      PC1541      PC1542
## 0.962839134 0.962955275 0.963071169 0.963186767 0.963301891 0.963416819
##      PC1543      PC1544      PC1545      PC1546      PC1547      PC1548
## 0.963531640 0.963646316 0.963760772 0.963875164 0.963989071 0.964102920
##      PC1549      PC1550      PC1551      PC1552      PC1553      PC1554
## 0.964216550 0.964329930 0.964442980 0.964555938 0.964668551 0.964780909
##      PC1555      PC1556      PC1557      PC1558      PC1559      PC1560
## 0.964893045 0.965004990 0.965116760 0.965228195 0.965339302 0.965450233
##      PC1561      PC1562      PC1563      PC1564      PC1565      PC1566
## 0.965561028 0.965671686 0.965782089 0.965892393 0.966002562 0.966112377
##      PC1567      PC1568      PC1569      PC1570      PC1571      PC1572
## 0.966221921 0.966331117 0.966439967 0.966548675 0.966657042 0.966765257
##      PC1573      PC1574      PC1575      PC1576      PC1577      PC1578
## 0.966873117 0.966980753 0.967088234 0.967195557 0.967302436 0.967409196
##      PC1579      PC1580      PC1581      PC1582      PC1583      PC1584
## 0.967515803 0.967622127 0.967728324 0.967834224 0.967939811 0.968045153
##      PC1585      PC1586      PC1587      PC1588      PC1589      PC1590
## 0.968150273 0.968255115 0.968359618 0.968464077 0.968568185 0.968672242
##      PC1591      PC1592      PC1593      PC1594      PC1595      PC1596
## 0.968775906 0.968879514 0.968982916 0.969085936 0.969188898 0.969291529
##      PC1597      PC1598      PC1599      PC1600      PC1601      PC1602
## 0.969393916 0.969496120 0.969597938 0.969699730 0.969801177 0.969902353
##      PC1603      PC1604      PC1605      PC1606      PC1607      PC1608
## 0.970003401 0.970104201 0.970204761 0.970305020 0.970405114 0.970504837
##      PC1609      PC1610      PC1611      PC1612      PC1613      PC1614
## 0.970604433 0.970703694 0.970802701 0.970901557 0.971000266 0.971098603
##      PC1615      PC1616      PC1617      PC1618      PC1619      PC1620
## 0.971196488 0.971294272 0.971391937 0.971489153 0.971586124 0.971683022
##      PC1621      PC1622      PC1623      PC1624      PC1625      PC1626
## 0.971779830 0.971876228 0.971972415 0.972068522 0.972164402 0.972260070
##      PC1627      PC1628      PC1629      PC1630      PC1631      PC1632
## 0.972355498 0.972450733 0.972545845 0.972640711 0.972735367 0.972829710
##      PC1633      PC1634      PC1635      PC1636      PC1637      PC1638
## 0.972923970 0.973017739 0.973111386 0.973204775 0.973297864 0.973390916
##      PC1639      PC1640      PC1641      PC1642      PC1643      PC1644
## 0.973483852 0.973576349 0.973668824 0.973760655 0.973852325 0.973943826
##      PC1645      PC1646      PC1647      PC1648      PC1649      PC1650
## 0.974035266 0.974126386 0.974217416 0.974308315 0.974398963 0.974489265
##      PC1651      PC1652      PC1653      PC1654      PC1655      PC1656
## 0.974579397 0.974669448 0.974759294 0.974848943 0.974938205 0.975027267
##      PC1657      PC1658      PC1659      PC1660      PC1661      PC1662
## 0.975116203 0.975204872 0.975293442 0.975381811 0.975469832 0.975557572
##      PC1663      PC1664      PC1665      PC1666      PC1667      PC1668
## 0.975645073 0.975732475 0.975819667 0.975906836 0.975993570 0.976080235
##      PC1669      PC1670      PC1671      PC1672      PC1673      PC1674
## 0.976166689 0.976252999 0.976338944 0.976424539 0.976510074 0.976595337
##      PC1675      PC1676      PC1677      PC1678      PC1679      PC1680
## 0.976680424 0.976765257 0.976849872 0.976934390 0.977018564 0.977102491
##      PC1681      PC1682      PC1683      PC1684      PC1685      PC1686
## 0.977186278 0.977270027 0.977353291 0.977436408 0.977519406 0.977602172
##      PC1687      PC1688      PC1689      PC1690      PC1691      PC1692
## 0.977684745 0.977767281 0.977849701 0.977931642 0.978013424 0.978094929
##      PC1693      PC1694      PC1695      PC1696      PC1697      PC1698
```

```
## 0.978176402 0.978257551 0.978338622 0.978419465 0.978499951 0.978580299
##      PC1699      PC1700      PC1701      PC1702      PC1703      PC1704
## 0.978660561 0.978740722 0.978820680 0.978900323 0.978979612 0.979058718
##      PC1705      PC1706      PC1707      PC1708      PC1709      PC1710
## 0.979137757 0.979216581 0.979295335 0.979373712 0.979451815 0.979529838
##      PC1711      PC1712      PC1713      PC1714      PC1715      PC1716
## 0.979607681 0.979685404 0.979762952 0.979840421 0.979917698 0.979994711
##      PC1717      PC1718      PC1719      PC1720      PC1721      PC1722
## 0.980071624 0.980148444 0.980224960 0.980301282 0.980377342 0.980453179
##      PC1723      PC1724      PC1725      PC1726      PC1727      PC1728
## 0.980528789 0.980604149 0.980679266 0.980754199 0.980829007 0.980903512
##      PC1729      PC1730      PC1731      PC1732      PC1733      PC1734
## 0.980977948 0.981051959 0.981125882 0.981199593 0.981273151 0.981346613
##      PC1735      PC1736      PC1737      PC1738      PC1739      PC1740
## 0.981419893 0.981493034 0.981566094 0.981639019 0.981711524 0.981783853
##      PC1741      PC1742      PC1743      PC1744      PC1745      PC1746
## 0.981856136 0.981928057 0.981999794 0.982071343 0.982142551 0.982213724
##      PC1747      PC1748      PC1749      PC1750      PC1751      PC1752
## 0.982284790 0.982355413 0.982425859 0.982496218 0.982566352 0.982636163
##      PC1753      PC1754      PC1755      PC1756      PC1757      PC1758
## 0.982705874 0.982775515 0.982845054 0.982914267 0.982983349 0.983052385
##      PC1759      PC1760      PC1761      PC1762      PC1763      PC1764
## 0.983121299 0.983189870 0.983258377 0.983326667 0.983394726 0.983462671
##      PC1765      PC1766      PC1767      PC1768      PC1769      PC1770
## 0.983530372 0.983597961 0.983665345 0.983732544 0.983799554 0.983866525
##      PC1771      PC1772      PC1773      PC1774      PC1775      PC1776
## 0.983933237 0.983999787 0.984065956 0.984132053 0.984197941 0.984263749
##      PC1777      PC1778      PC1779      PC1780      PC1781      PC1782
## 0.984329295 0.984394553 0.984459734 0.984524741 0.984589505 0.984654084
##      PC1783      PC1784      PC1785      PC1786      PC1787      PC1788
## 0.984718399 0.984782492 0.984846510 0.984910273 0.984973961 0.985037468
##      PC1789      PC1790      PC1791      PC1792      PC1793      PC1794
## 0.985100768 0.985163927 0.985226887 0.985289767 0.985352281 0.985414516
##      PC1795      PC1796      PC1797      PC1798      PC1799      PC1800
## 0.985476533 0.985538493 0.985600162 0.985661773 0.985723097 0.985784333
##      PC1801      PC1802      PC1803      PC1804      PC1805      PC1806
## 0.985845442 0.985906438 0.985967266 0.986027882 0.986088241 0.986148511
##      PC1807      PC1808      PC1809      PC1810      PC1811      PC1812
## 0.986208630 0.986268582 0.986328153 0.986387667 0.986447150 0.986506386
##      PC1813      PC1814      PC1815      PC1816      PC1817      PC1818
## 0.986565488 0.986624466 0.986683305 0.986742048 0.986800601 0.986858972
##      PC1819      PC1820      PC1821      PC1822      PC1823      PC1824
## 0.986917111 0.986975138 0.987032894 0.987090547 0.987148008 0.987205271
##      PC1825      PC1826      PC1827      PC1828      PC1829      PC1830
## 0.987262454 0.987319585 0.987376608 0.987433443 0.987489931 0.987546241
##      PC1831      PC1832      PC1833      PC1834      PC1835      PC1836
## 0.987602446 0.987658419 0.987714316 0.987769884 0.987825329 0.987880576
##      PC1837      PC1838      PC1839      PC1840      PC1841      PC1842
## 0.987935588 0.987990391 0.988045098 0.988099657 0.988154058 0.988208319
##      PC1843      PC1844      PC1845      PC1846      PC1847      PC1848
## 0.988262396 0.988316365 0.988370130 0.988423770 0.988477212 0.988530498
##      PC1849      PC1850      PC1851      PC1852      PC1853      PC1854
## 0.988583552 0.988636538 0.988689414 0.988742041 0.988794549 0.988846928
##      PC1855      PC1856      PC1857      PC1858      PC1859      PC1860
```

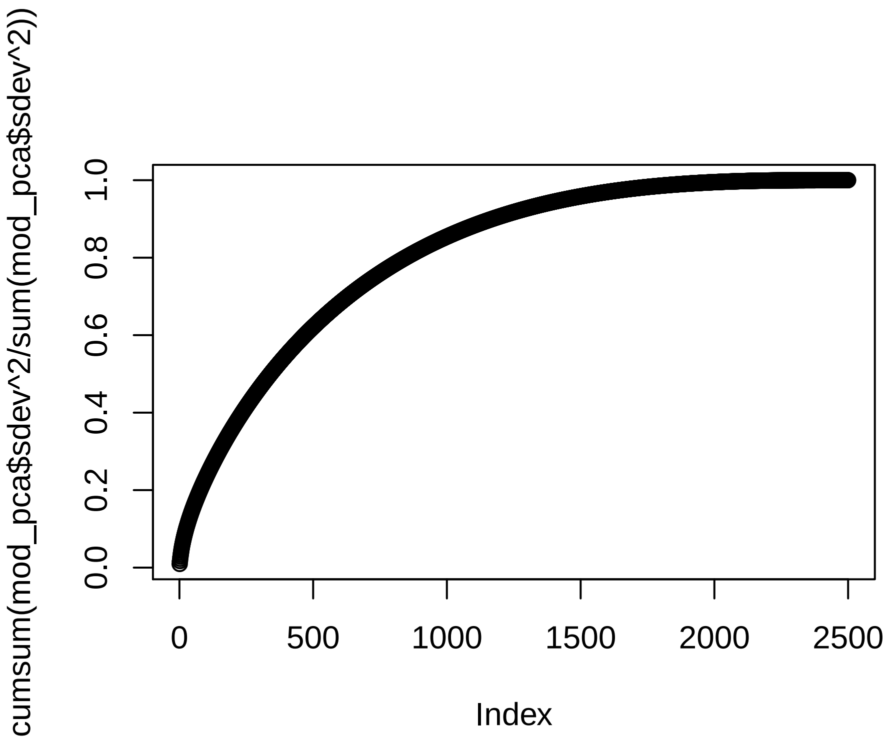
```
## 0.988899101 0.988951158 0.989003146 0.989054918 0.989106410 0.989157766
##      PC1861      PC1862      PC1863      PC1864      PC1865      PC1866
## 0.989209005 0.989260131 0.989311059 0.989361900 0.989412695 0.989463320
##      PC1867      PC1868      PC1869      PC1870      PC1871      PC1872
## 0.989513755 0.989564047 0.989614062 0.989664031 0.989713960 0.989763759
##      PC1873      PC1874      PC1875      PC1876      PC1877      PC1878
## 0.989813307 0.989862800 0.989912130 0.989961352 0.990010359 0.990059317
##      PC1879      PC1880      PC1881      PC1882      PC1883      PC1884
## 0.990108076 0.990156694 0.990205150 0.990253382 0.990301483 0.990349358
##      PC1885      PC1886      PC1887      PC1888      PC1889      PC1890
## 0.990397087 0.990444667 0.990492123 0.990539270 0.990586357 0.990633146
##      PC1891      PC1892      PC1893      PC1894      PC1895      PC1896
## 0.990679831 0.990726400 0.990772865 0.990819248 0.990865405 0.990911472
##      PC1897      PC1898      PC1899      PC1900      PC1901      PC1902
## 0.990957340 0.991002982 0.991048470 0.991093850 0.991139115 0.991184277
##      PC1903      PC1904      PC1905      PC1906      PC1907      PC1908
## 0.991229283 0.991274099 0.991318694 0.991363185 0.991407582 0.991451908
##      PC1909      PC1910      PC1911      PC1912      PC1913      PC1914
## 0.991496110 0.991540027 0.991583756 0.991627299 0.991670810 0.991714171
##      PC1915      PC1916      PC1917      PC1918      PC1919      PC1920
## 0.991757488 0.991800538 0.991843509 0.991886248 0.991928923 0.991971391
##      PC1921      PC1922      PC1923      PC1924      PC1925      PC1926
## 0.992013709 0.992055872 0.992097792 0.992139452 0.992181034 0.992222557
##      PC1927      PC1928      PC1929      PC1930      PC1931      PC1932
## 0.992263983 0.992305176 0.992346272 0.992387188 0.992428044 0.992468699
##      PC1933      PC1934      PC1935      PC1936      PC1937      PC1938
## 0.992509287 0.992549639 0.992589937 0.992630066 0.992670134 0.992709977
##      PC1939      PC1940      PC1941      PC1942      PC1943      PC1944
## 0.992749717 0.992789396 0.992828912 0.992868348 0.992907686 0.992946786
##      PC1945      PC1946      PC1947      PC1948      PC1949      PC1950
## 0.992985824 0.993024684 0.993063444 0.993102074 0.993140481 0.993178829
##      PC1951      PC1952      PC1953      PC1954      PC1955      PC1956
## 0.993216792 0.993254700 0.993292502 0.993330132 0.993367621 0.993404958
##      PC1957      PC1958      PC1959      PC1960      PC1961      PC1962
## 0.993442241 0.993479378 0.993516422 0.993553396 0.993590156 0.993626711
##      PC1963      PC1964      PC1965      PC1966      PC1967      PC1968
## 0.993663086 0.993699362 0.993735505 0.993771596 0.993807630 0.993843471
##      PC1969      PC1970      PC1971      PC1972      PC1973      PC1974
## 0.993879192 0.993914746 0.993950227 0.993985554 0.994020813 0.994055991
##      PC1975      PC1976      PC1977      PC1978      PC1979      PC1980
## 0.994090971 0.994125849 0.994160558 0.994195128 0.994229564 0.994263880
##      PC1981      PC1982      PC1983      PC1984      PC1985      PC1986
## 0.994298123 0.994332166 0.994366153 0.994399952 0.994433547 0.994466983
##      PC1987      PC1988      PC1989      PC1990      PC1991      PC1992
## 0.994500336 0.994533669 0.994566890 0.994599982 0.994632979 0.994665815
##      PC1993      PC1994      PC1995      PC1996      PC1997      PC1998
## 0.994698513 0.994731091 0.994763598 0.994795958 0.994828044 0.994860055
##      PC1999      PC2000      PC2001      PC2002      PC2003      PC2004
## 0.994891939 0.994923657 0.994955244 0.994986749 0.995018109 0.995049382
##      PC2005      PC2006      PC2007      PC2008      PC2009      PC2010
## 0.995080621 0.995111496 0.995142275 0.995172974 0.995203531 0.995234028
##      PC2011      PC2012      PC2013      PC2014      PC2015      PC2016
## 0.995264381 0.995294649 0.995324838 0.995354991 0.995384992 0.995414782
##      PC2017      PC2018      PC2019      PC2020      PC2021      PC2022
```

```
## 0.995444557 0.995474184 0.995503735 0.995533067 0.995562352 0.995591495
##      PC2023      PC2024      PC2025      PC2026      PC2027      PC2028
## 0.995620348 0.995649162 0.995677841 0.995706490 0.995735098 0.995763658
##      PC2029      PC2030      PC2031      PC2032      PC2033      PC2034
## 0.995791971 0.995820088 0.995848107 0.995876036 0.995903829 0.995931574
##      PC2035      PC2036      PC2037      PC2038      PC2039      PC2040
## 0.995959300 0.995986667 0.996013969 0.996041182 0.996068359 0.996095368
##      PC2041      PC2042      PC2043      PC2044      PC2045      PC2046
## 0.996122298 0.996149056 0.996175668 0.996202178 0.996228639 0.996254961
##      PC2047      PC2048      PC2049      PC2050      PC2051      PC2052
## 0.996281097 0.996307145 0.996333125 0.996359050 0.996384681 0.996410048
##      PC2053      PC2054      PC2055      PC2056      PC2057      PC2058
## 0.996435389 0.996460673 0.996485832 0.996510855 0.996535837 0.996560725
##      PC2059      PC2060      PC2061      PC2062      PC2063      PC2064
## 0.996585493 0.996610088 0.996634607 0.996658919 0.996683136 0.996707242
##      PC2065      PC2066      PC2067      PC2068      PC2069      PC2070
## 0.996731341 0.996755367 0.996779255 0.996802925 0.996826526 0.996850015
##      PC2071      PC2072      PC2073      PC2074      PC2075      PC2076
## 0.996873435 0.996896730 0.996919873 0.996942975 0.996965897 0.996988657
##      PC2077      PC2078      PC2079      PC2080      PC2081      PC2082
## 0.997011402 0.997033958 0.997056472 0.997078892 0.997101240 0.997123413
##      PC2083      PC2084      PC2085      PC2086      PC2087      PC2088
## 0.997145478 0.997167405 0.997189184 0.997210858 0.997232458 0.997254033
##      PC2089      PC2090      PC2091      PC2092      PC2093      PC2094
## 0.997275477 0.997296797 0.997318046 0.997339112 0.997360109 0.997380978
##      PC2095      PC2096      PC2097      PC2098      PC2099      PC2100
## 0.997401754 0.997422412 0.997443039 0.997463609 0.997484107 0.997504434
##      PC2101      PC2102      PC2103      PC2104      PC2105      PC2106
## 0.997524650 0.997544781 0.997564800 0.997584709 0.997604492 0.997624130
##      PC2107      PC2108      PC2109      PC2110      PC2111      PC2112
## 0.997643704 0.997663182 0.997682559 0.997701852 0.997721042 0.997740117
##      PC2113      PC2114      PC2115      PC2116      PC2117      PC2118
## 0.997759144 0.997778039 0.997796851 0.997815573 0.997834113 0.997852623
##      PC2119      PC2120      PC2121      PC2122      PC2123      PC2124
## 0.997871012 0.997889370 0.997907544 0.997925611 0.997943607 0.997961526
##      PC2125      PC2126      PC2127      PC2128      PC2129      PC2130
## 0.997979360 0.997997080 0.998014715 0.998032300 0.998049841 0.998067173
##      PC2131      PC2132      PC2133      PC2134      PC2135      PC2136
## 0.998084322 0.998101431 0.998118420 0.998135324 0.998152165 0.998168951
##      PC2137      PC2138      PC2139      PC2140      PC2141      PC2142
## 0.998185644 0.998202200 0.998218602 0.998234918 0.998251211 0.998267387
##      PC2143      PC2144      PC2145      PC2146      PC2147      PC2148
## 0.998283429 0.998299445 0.998315337 0.998331197 0.998346903 0.998362543
##      PC2149      PC2150      PC2151      PC2152      PC2153      PC2154
## 0.998378133 0.998393590 0.998408989 0.998424314 0.998439439 0.998454536
##      PC2155      PC2156      PC2157      PC2158      PC2159      PC2160
## 0.998469435 0.998484273 0.998499086 0.998513818 0.998528491 0.998543004
##      PC2161      PC2162      PC2163      PC2164      PC2165      PC2166
## 0.998557401 0.998571744 0.998585961 0.998600090 0.998614194 0.998628230
##      PC2167      PC2168      PC2169      PC2170      PC2171      PC2172
## 0.998642197 0.998656003 0.998669692 0.998683273 0.998696822 0.998710271
##      PC2173      PC2174      PC2175      PC2176      PC2177      PC2178
## 0.998723440 0.998736548 0.998749631 0.998762618 0.998775516 0.998788324
##      PC2179      PC2180      PC2181      PC2182      PC2183      PC2184
```

```
## 0.998801054 0.998813750 0.998826344 0.998838782 0.998851198 0.998863533
##      PC2185      PC2186      PC2187      PC2188      PC2189      PC2190
## 0.998875787 0.998887999 0.998900117 0.998912203 0.998924048 0.998935842
##      PC2191      PC2192      PC2193      PC2194      PC2195      PC2196
## 0.998947546 0.998959224 0.998970735 0.998982174 0.998993593 0.999004878
##      PC2197      PC2198      PC2199      PC2200      PC2201      PC2202
## 0.999016133 0.999027286 0.999038342 0.999049323 0.999060234 0.999071128
##      PC2203      PC2204      PC2205      PC2206      PC2207      PC2208
## 0.999081846 0.999092476 0.999103060 0.999113539 0.999123935 0.999134263
##      PC2209      PC2210      PC2211      PC2212      PC2213      PC2214
## 0.999144546 0.999154794 0.999164914 0.999174879 0.999184758 0.999194619
##      PC2215      PC2216      PC2217      PC2218      PC2219      PC2220
## 0.999204393 0.999214085 0.999223766 0.999233341 0.999242853 0.999252301
##      PC2221      PC2222      PC2223      PC2224      PC2225      PC2226
## 0.999261664 0.999270884 0.999280039 0.999289125 0.999298118 0.999307062
##      PC2227      PC2228      PC2229      PC2230      PC2231      PC2232
## 0.999315955 0.999324779 0.999333498 0.999342120 0.999350604 0.999359068
##      PC2233      PC2234      PC2235      PC2236      PC2237      PC2238
## 0.999367440 0.999375734 0.999383977 0.999392122 0.999400197 0.999408262
##      PC2239      PC2240      PC2241      PC2242      PC2243      PC2244
## 0.999416255 0.999424121 0.999431925 0.999439702 0.999447399 0.999455047
##      PC2245      PC2246      PC2247      PC2248      PC2249      PC2250
## 0.999462620 0.999470137 0.999477581 0.999484945 0.999492278 0.999499507
##      PC2251      PC2252      PC2253      PC2254      PC2255      PC2256
## 0.999506666 0.999513774 0.999520798 0.999527740 0.999534591 0.999541392
##      PC2257      PC2258      PC2259      PC2260      PC2261      PC2262
## 0.999548082 0.999554714 0.999561264 0.999567750 0.999574190 0.999580585
##      PC2263      PC2264      PC2265      PC2266      PC2267      PC2268
## 0.999586949 0.999593210 0.999599420 0.999605579 0.999611651 0.999617685
##      PC2269      PC2270      PC2271      PC2272      PC2273      PC2274
## 0.999623701 0.999629629 0.999635446 0.999641228 0.999646935 0.999652585
##      PC2275      PC2276      PC2277      PC2278      PC2279      PC2280
## 0.999658155 0.999663657 0.999669118 0.999674560 0.999679982 0.999685291
##      PC2281      PC2282      PC2283      PC2284      PC2285      PC2286
## 0.999690458 0.999695557 0.999700612 0.999705649 0.999710602 0.999715535
##      PC2287      PC2288      PC2289      PC2290      PC2291      PC2292
## 0.999720334 0.999725120 0.999729839 0.999734521 0.999739120 0.999743648
##      PC2293      PC2294      PC2295      PC2296      PC2297      PC2298
## 0.999748164 0.999752600 0.999756969 0.999761319 0.999765590 0.999769786
##      PC2299      PC2300      PC2301      PC2302      PC2303      PC2304
## 0.999773940 0.999778018 0.999782059 0.999786055 0.999790036 0.999793929
##      PC2305      PC2306      PC2307      PC2308      PC2309      PC2310
## 0.999797803 0.999801581 0.999805341 0.999809079 0.999812767 0.999816381
##      PC2311      PC2312      PC2313      PC2314      PC2315      PC2316
## 0.999819939 0.999823436 0.999826882 0.999830280 0.999833641 0.999836939
##      PC2317      PC2318      PC2319      PC2320      PC2321      PC2322
## 0.999840198 0.999843435 0.999846620 0.999849764 0.999852849 0.999855891
##      PC2323      PC2324      PC2325      PC2326      PC2327      PC2328
## 0.999858875 0.999861803 0.999864702 0.999867526 0.999870329 0.999873060
##      PC2329      PC2330      PC2331      PC2332      PC2333      PC2334
## 0.999875768 0.999878467 0.999881116 0.999883704 0.999886261 0.999888774
##      PC2335      PC2336      PC2337      PC2338      PC2339      PC2340
## 0.999891266 0.999893703 0.999896109 0.999898484 0.999900794 0.999903094
##      PC2341      PC2342      PC2343      PC2344      PC2345      PC2346
```

```
## 0.999905339 0.999907545 0.999909712 0.999911835 0.999913909 0.999915967
##      PC2347      PC2348      PC2349      PC2350      PC2351      PC2352
## 0.999917994 0.999920011 0.999921997 0.999923956 0.999925864 0.999927755
##      PC2353      PC2354      PC2355      PC2356      PC2357      PC2358
## 0.999929607 0.999931426 0.999933191 0.999934928 0.999936657 0.999938332
##      PC2359      PC2360      PC2361      PC2362      PC2363      PC2364
## 0.999939989 0.999941626 0.999943229 0.999944797 0.999946352 0.999947872
##      PC2365      PC2366      PC2367      PC2368      PC2369      PC2370
## 0.999949360 0.999950806 0.999952243 0.999953653 0.999955037 0.999956391
##      PC2371      PC2372      PC2373      PC2374      PC2375      PC2376
## 0.999957709 0.999959022 0.999960321 0.999961608 0.999962863 0.999964103
##      PC2377      PC2378      PC2379      PC2380      PC2381      PC2382
## 0.999965317 0.999966503 0.999967658 0.999968812 0.999969923 0.999971017
##      PC2383      PC2384      PC2385      PC2386      PC2387      PC2388
## 0.999972109 0.999973166 0.999974199 0.999975184 0.999976162 0.999977116
##      PC2389      PC2390      PC2391      PC2392      PC2393      PC2394
## 0.999978031 0.999978915 0.999979785 0.999980623 0.999981448 0.999982259
##      PC2395      PC2396      PC2397      PC2398      PC2399      PC2400
## 0.999983060 0.999983836 0.999984587 0.999985311 0.999986008 0.999986690
##      PC2401      PC2402      PC2403      PC2404      PC2405      PC2406
## 0.999987346 0.999987975 0.999988589 0.999989193 0.999989760 0.999990312
##      PC2407      PC2408      PC2409      PC2410      PC2411      PC2412
## 0.999990858 0.999991388 0.999991905 0.999992416 0.999992913 0.999993385
##      PC2413      PC2414      PC2415      PC2416      PC2417      PC2418
## 0.999993828 0.999994267 0.999994695 0.999995108 0.999995508 0.999995888
##      PC2419      PC2420      PC2421      PC2422      PC2423      PC2424
## 0.999996231 0.999996540 0.999996830 0.999997113 0.999997382 0.999997626
##      PC2425      PC2426      PC2427      PC2428      PC2429      PC2430
## 0.999997858 0.999998079 0.999998278 0.999998474 0.999998654 0.999998827
##      PC2431      PC2432      PC2433      PC2434      PC2435      PC2436
## 0.999998951 0.999999071 0.999999177 0.999999281 0.999999375 0.999999465
##      PC2437      PC2438      PC2439      PC2440      PC2441      PC2442
## 0.999999549 0.999999624 0.999999696 0.999999755 0.999999805 0.999999846
##      PC2443      PC2444      PC2445      PC2446      PC2447      PC2448
## 0.999999877 0.999999905 0.999999930 0.999999952 0.999999967 0.999999977
##      PC2449      PC2450      PC2451      PC2452      PC2453      PC2454
## 0.999999986 0.999999995 0.999999999 1.000000000 1.000000000 1.000000000
##      PC2455      PC2456      PC2457      PC2458      PC2459      PC2460
## 1.000000000 1.000000000 1.000000000 1.000000000 1.000000000 1.000000000
##      PC2461      PC2462      PC2463      PC2464      PC2465      PC2466
## 1.000000000 1.000000000 1.000000000 1.000000000 1.000000000 1.000000000
##      PC2467      PC2468      PC2469      PC2470      PC2471      PC2472
## 1.000000000 1.000000000 1.000000000 1.000000000 1.000000000 1.000000000
##      PC2473      PC2474      PC2475      PC2476      PC2477      PC2478
## 1.000000000 1.000000000 1.000000000 1.000000000 1.000000000 1.000000000
##      PC2479      PC2480      PC2481      PC2482      PC2483      PC2484
## 1.000000000 1.000000000 1.000000000 1.000000000 1.000000000 1.000000000
##      PC2485      PC2486      PC2487      PC2488      PC2489      PC2490
## 1.000000000 1.000000000 1.000000000 1.000000000 1.000000000 1.000000000
##      PC2491      PC2492      PC2493      PC2494      PC2495      PC2496
## 1.000000000 1.000000000 1.000000000 1.000000000 1.000000000 1.000000000
##      PC2497      PC2498      PC2499      PC2500
## 1.000000000 1.000000000 1.000000000 1.000000000
```

```r
plot(cumsum(mod_pca$sdev^2/sum(mod_pca$sdev^2)))
```




## Step 4:

Now that we have run the principal component analysis, we need to wrangle the data so that we can fit some proper classification models.

```r
train_class = data.frame(mod_pca$x[,1:730])
train_class['author']=labels
train_load = mod_pca$rotation[,1:730]
test_class_pre <- scale(DTM_test1) %*% train_load
test_class <- as.data.frame(test_class_pre)
test_class['author']=labels1
```

## Step 5:

Now that we have classification-ready training and test sets, and a sufficient principal component analysis that seeks to explain a large portion of the data, we can fit a tree model, specifically aggregating tree models using random forests. This way we can seek to reduce out of sample error as much as possible.

we then seek to test the predictive power of our random forest model. We see that

```r
set.seed(1)
rf_model<-randomForest(as.factor(author)~.,data=train_class, mtry=6,importance=TRUE)

rf_predict<-predict(rf_model,data=test_class)
rf_table<-as.data.frame(table(rf_predict,as.factor(test_class$author)))

predicted<-rf_predict
```

```r
actual<-as.factor(test_class$author)

temp<-as.data.frame(cbind(actual,predicted))
temp$flag<-ifelse(temp$actual==temp$predicted,1,0)
sum(temp$flag)
```

```
## [1] 1881
```

```r
sum(temp$flag)*100/nrow(temp)
```

```
## [1] 75.24
```

## Step 6:

Now that we have a random forest model, we need to run a second model. We chose to run a KNN classification, in large part because of the flexibility of a KNN model. We need to ensure that we don't overfit when choosing a K, but we were interested to see how the improved flexibility over a tree model might improve our out of sample error.

```r
train.X = subset(train_class, select = -c(author))
test.X = subset(test_class,select=-c(author))
train.author=as.factor(train_class$author)
test.author=as.factor(test_class$author)

set.seed(1)
knn_pred=knn(train.X,test.X,train.author,k=5)

temp_knn=as.data.frame(cbind(knn_pred,test.author))
temp_knn_flag<-ifelse(as.integer(knn_pred)==as.integer(test.author),1,0)
sum(temp_knn_flag)
```

```
## [1] 828
```

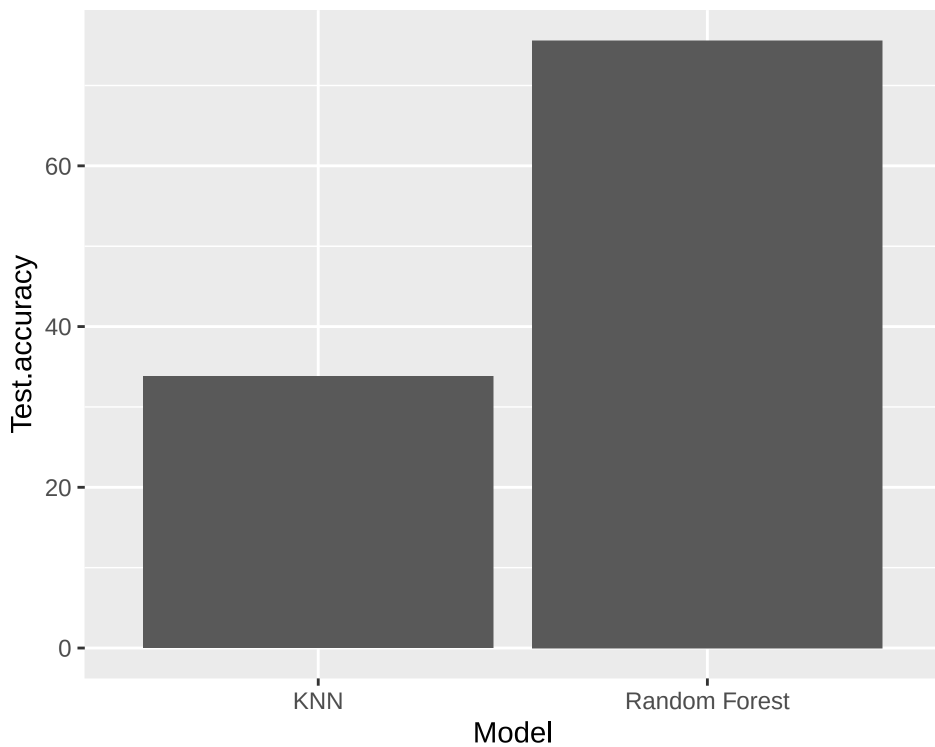
```r
sum(temp_knn_flag)*100/nrow(temp_knn)
```

```
## [1] 33.12
```

```r
comp<-data.frame("Model"=c("Random Forest","KNN"), "Test.accuracy"=c(75.6,33.84))
comp
```

```
##           Model Test.accuracy
## 1 Random Forest         75.60
## 2           KNN         33.84
```

```r
ggplot(comp,aes(x=Model,y=Test.accuracy))+geom_col()
```




# Results:

Clearly we were wrong about our KNN model. The random forest, despite the relative rigidity of tree models, did much better than our KNN model. This is likely thanks to the aggregation of the trees which gives our random forest the ability to classify based on many more variations of the model than a KNN has access to.

It's possible that we could improve upon the KNN using various methods of finding the optimal K, we did alter the K used for the model and found that K=5 gave us the best out of sample error but usually K=5 when dealing with a dataset of this size is an example of overfitting. Because of this we are not confident that the KNN model is the best model to use.

## Question 6 - Association Rule Mining

Here we look at a data set containing 9835 different baskets of groceries purchased while shopping to hopefully discover some insightful associations between groceries in these baskets.

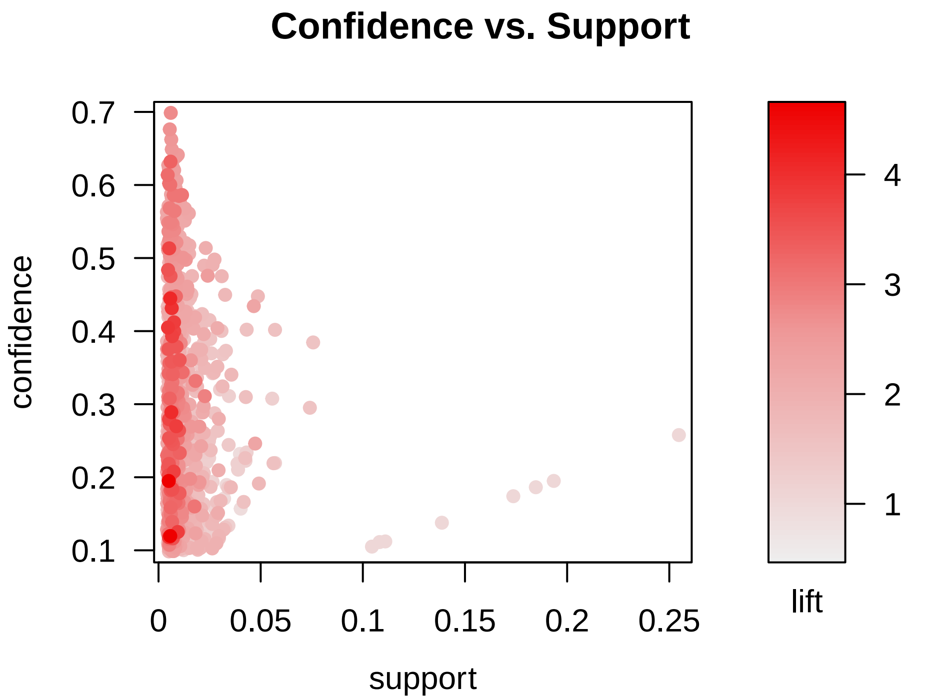
After cleaning the data, we can run it through an apriori algorithm to evaluate association rules between individual grocery items. Apriori returns 1582 total rules.

Now we can look at some of individual the association rules. We are most interested in associations with a relatively high confidence (how often one food item appears in transactions that contain another given item(s)), and high lift (measure of the chances a customer will buy one food item if the customer has already bought another given item(s)).

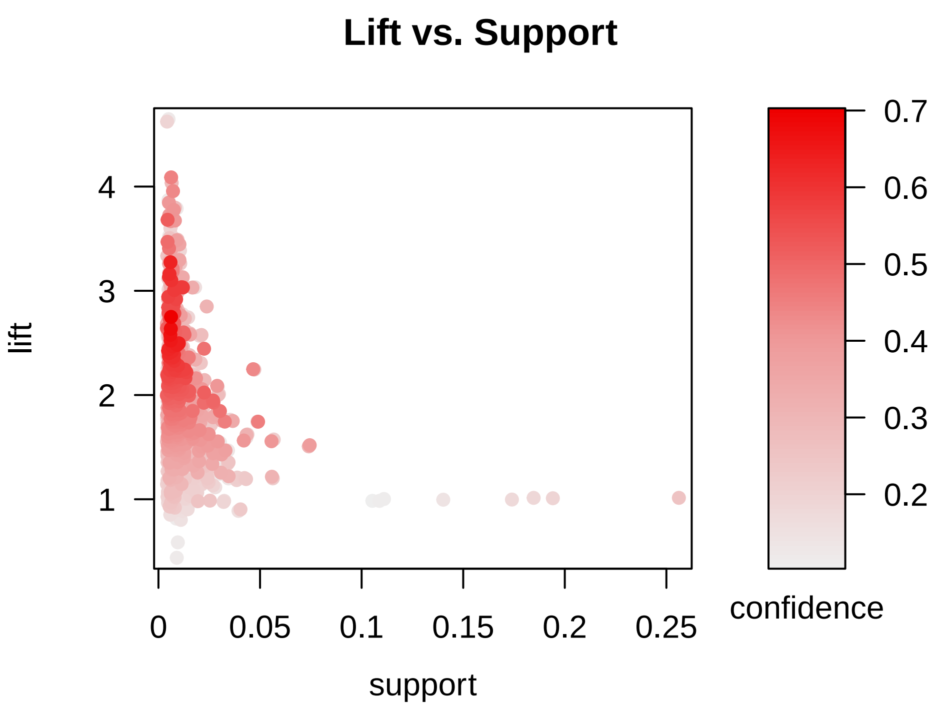
A lift greater than 1 implies an increase in purchase probability between two sets of items. The maximum lift of all the rules we've generated is 4.63, with the maximum confidence being .7 so we will want to look at a subset in that upper range. We can look at a subset of rules with confidence greater than .45 and lift greater than 3. 10 rules fall into this subset:

These association rules generally make intuitive sense. Vegetables are associated with purchases of other vegetables, fruit with yogurt, etc. We also notice that some of these categories are quite vague and thus might be more widely associated with many other categories (i.e. "other vegetables").

We can plot the rules together in a Confidence vs. Support as well as a Lift vs. Support plot:




```
## To reduce overplotting, jitter is added! Use jitter = 0 to prevent jitter.
```




From these plots, we can see a subset of the association rules that have a relatively high support but low lift and confidence. We can look at this subset by limiting our rules to those with a support greater than 0.04.

They are shown below:

The rules with high support (fraction of grocery carts that contain both items on the left and on the right) tend to involve items that are extremely popular in general (milk, yogurt, other vegetables). This makes sense as to why they frequently have high support but low lift.

Finally, we can visualize a network of the association rules generated by the apriori algorithm. From this magnificent plot we can infer that there are a few central items that are closely related to almost everything else in the grocery store including: whole milk, other vegetables, yogurt, and root vegetables. We also see some items of secondary importance such as sausage, whipped cream, and eggs.

All in all these rules make very much intuitive sense when we consider items that are traditionally popular at grocery stores.