

STAT139 Final Project

```
winedata = read.csv('winemag-data-130k-v2.csv')

newwinedata = read.csv('new_wine_data.csv')

# create "Other" country
country_cat = vector()
for (i in 1:nrow(winedata)) {
  if (winedata$country[i] == 'Italy') {
    country_cat = c(country_cat, 'Italy')
  }
  else if (winedata$country[i] == 'US') {
    country_cat = c(country_cat, 'US')
  }
  else if (winedata$country[i] == 'France') {
    country_cat = c(country_cat, 'France')
  }
  else if (winedata$country[i] == 'Spain') {
    country_cat = c(country_cat, 'Spain')
  }
  else if (winedata$country[i] == 'Portugal') {
    country_cat = c(country_cat, 'Portugal')
  }
  else if (winedata$country[i] == 'Austria') {
    country_cat = c(country_cat, 'Austria')
  }
  else if (winedata$country[i] == 'Chile') {
    country_cat = c(country_cat, 'Chile')
  }
  else if (winedata$country[i] == 'Argentina') {
    country_cat = c(country_cat, 'Argentina')
  }
  else if (winedata$country[i] == 'Australia') {
    country_cat = c(country_cat, 'Australia')
  }
  else if (winedata$country[i] == 'New Zealand') {
    country_cat = c(country_cat, 'New Zealand')
  }
  else if (winedata$country[i] == 'Germany') {
    country_cat = c(country_cat, 'Germany')
  }
  else {
    country_cat = c(country_cat, 'Other')
  }
}

# extract year
cleanyear = vector()
for (i in 1:nrow(newwinedata)) {
  if (is.na(newwinedata$title[i])) {
    cleanyear = c(cleanyear, NA)
  }
}
```

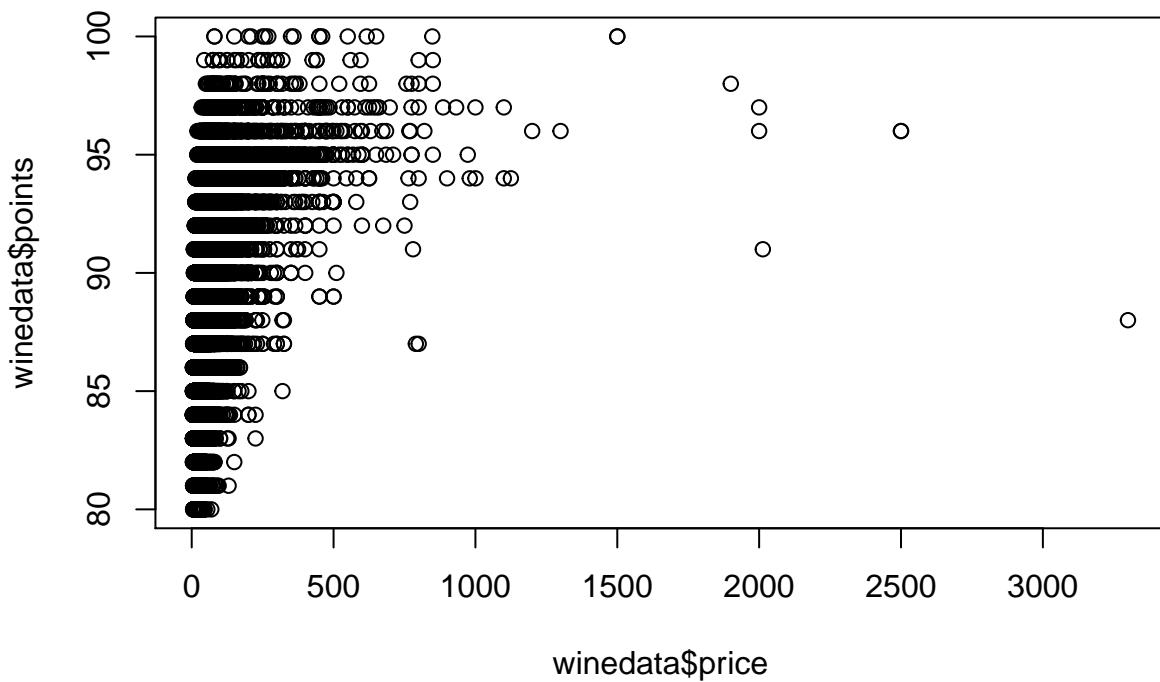
```

else if (as.numeric(newwinedata$title[i]) < 1700) {
  cleanyear = c(cleanyear,NA)
}
else if (as.numeric(newwinedata$title[i]) > 2018) {
  cleanyear = c(cleanyear,NA)
}
else {
  cleanyear = c(cleanyear,as.numeric(newwinedata$title[i]))
}
}

# add columns to dataframe
winedata$countrycat = country_cat
winedata$year = cleanyear

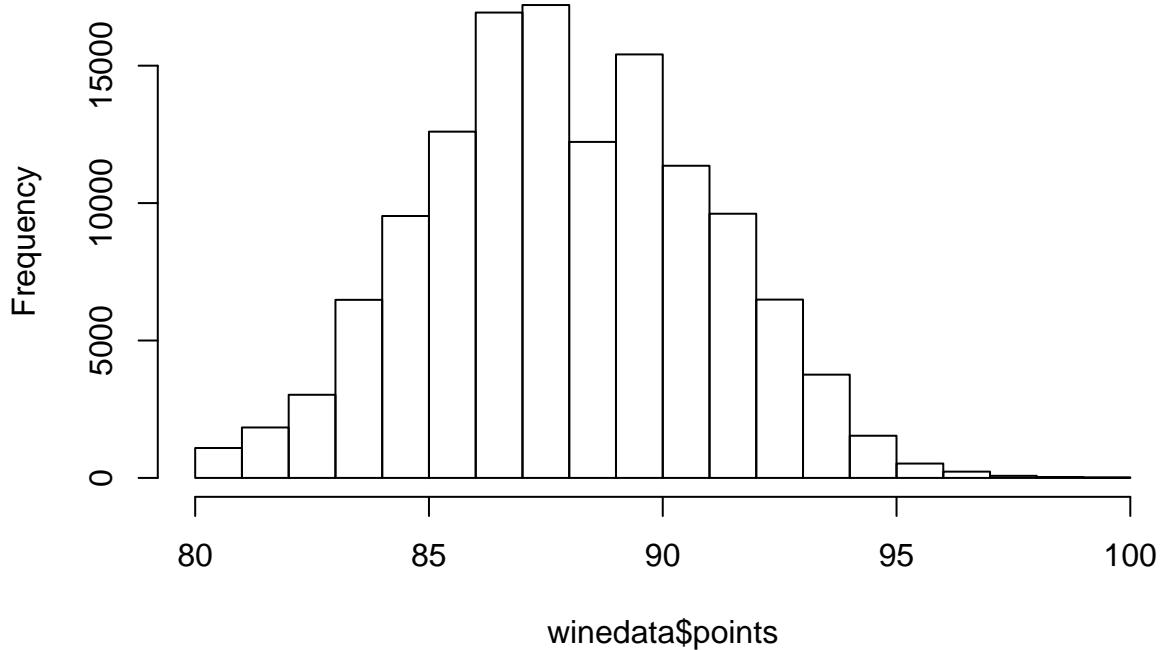
plot(winedata$price,winedata$points)

```

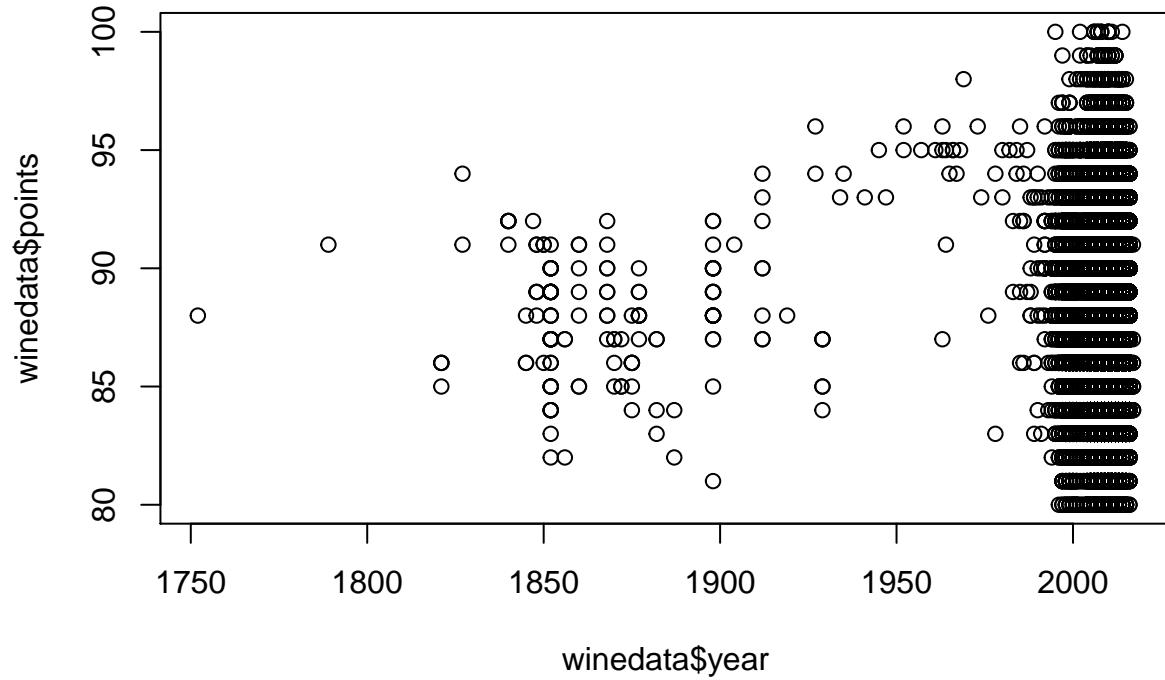


```
hist(winedata$points)
```

Histogram of winedata\$points



```
plot(winedata$year, winedata$points)
```



```
model1 = lm(points ~ countrycat + year + taster_name + log(price), data=winedata)
summary(model1)
```

```
##  
## Call:  
## lm(formula = points ~ countrycat + year + taster_name + log(price),
```

```

##      data = winedata)
##
## Residuals:
##      Min      1Q   Median      3Q     Max 
## -14.4215 -1.4630  0.0909  1.5649 10.4312 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                50.828302  2.320632 21.903 < 2e-16 ***
## countrycatAustralia        0.281028  0.125246  2.244 0.024847 *  
## countrycatAustria          0.305486  0.131015  2.332 0.019719 *  
## countrycatChile            0.154049  0.051157  3.011 0.002602 ** 
## countrycatFrance           -0.139440  0.118514 -1.177 0.239369  
## countrycatGermany          0.531496  0.121835  4.362 1.29e-05 *** 
## countrycatItaly             -0.046988 0.109325 -0.430 0.667341  
## countrycatNew Zealand      0.194932  0.132226  1.474 0.140422  
## countrycatOther             -0.364104 0.109741 -3.318 0.000907 *** 
## countrycatPortugal          0.662346  0.124217  5.332 9.72e-08 *** 
## countrycatSpain             0.312725  0.047727  6.552 5.69e-11 *** 
## countrycatUS                -0.993873 0.107064 -9.283 < 2e-16 *** 
## year                        0.013916  0.001153 12.069 < 2e-16 *** 
## taster_nameAlexander Peartree -1.376443 0.116927 -11.772 < 2e-16 *** 
## taster_nameAnna Lee C. Iijima  0.855351  0.046617 18.349 < 2e-16 *** 
## taster_nameAnne Krebiehl MW   2.182705  0.086550 25.219 < 2e-16 *** 
## taster_nameCarrie Dykes      -1.145094 0.199357 -5.744 9.27e-09 *** 
## taster_nameChristina Pickard  0.076346  1.023614  0.075 0.940546  
## taster_nameFiona Adams       -0.706658 0.488242 -1.447 0.147801  
## taster_nameJeff Jenssen      1.618690  0.126456 12.800 < 2e-16 *** 
## taster_nameJim Gordon        1.770568  0.040797 43.399 < 2e-16 *** 
## taster_nameJoe Czerwinski    0.249745  0.066152  3.775 0.000160 *** 
## taster_nameKerin O'Keefe      0.077051  0.037292  2.066 0.038813 *  
## taster_nameLauren Buzzeo     0.449782  0.081275  5.534 3.14e-08 *** 
## taster_nameMatt Kettmann     2.000410  0.034615 57.791 < 2e-16 *** 
## taster_nameMichael Schachner -0.587518 0.099994 -5.876 4.23e-09 *** 
## taster_nameMike DeSimone     1.440790  0.123955 11.623 < 2e-16 *** 
## taster_namePaul Gregutt      1.556635  0.029546 52.685 < 2e-16 *** 
## taster_nameRoger Voss        0.416837  0.071527  5.828 5.64e-09 *** 
## taster_nameSean P. Sullivan  1.191164  0.037651 31.637 < 2e-16 *** 
## taster_nameSusan Kostrzewska -0.206456  0.090705 -2.276 0.022840 *  
## taster_nameVirginie Boone    0.837132  0.030110 27.802 < 2e-16 *** 
## log(price)                  2.876298  0.010898 263.928 < 2e-16 *** 
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
##
## Residual standard error: 2.288 on 116775 degrees of freedom
##   (13163 observations deleted due to missingness)
## Multiple R-squared:  0.4354, Adjusted R-squared:  0.4352 
## F-statistic: 2814 on 32 and 116775 DF, p-value: < 2.2e-16

```