DATA 606: Project: Gold vs Silver Price

Anjal Hussan

Topic: Correlation between Gold and Silver Prices

Introduction:

Silver and gold are close substitutes for one another. Both have been used as currency in the pas. There is significant evidence that these metals being an attractive investment and can play a useful role in diversifying risk. There are also economic fundamentals that may act to drive the prices of silver and gold apart.

Research question:

The main objective of this project is to study the relationship and correlation between prices of silver and gold in commodity market.

Data Source:

The datasets for this project are from the following sites:

1. https://www.quandl.com/data/LBMA/GOLD-Gold-Price-London-Fixing (https://www.quandl.com/data/LBMA/GOLD-Gold-Price-London-Fixing)
2. https://www.quandl.com/data/LBMA/SILVER-Silver-Price-London-Fixing (https://www.quandl.com/data/LBMA/SILVER-Silver-Price-London-Fixing)

Cases:

Each case represents the price of silver and gold for a day. The full dataset represents data for 48 years with approximately 12618 cases.

Data Collection:

The data is collected from quandl.com

Type of study:

This is observational study.

Response:

Price of gold is the response variable. It is numerical continuous variable.

Explanatory:

The explanatory variable is the price of silver and is numerical.

Setup

```
#install.packages("tidyr")
#install.packages("magrittr")
#install.packages("sqldf")
#install.packages("ggplot2")
#install.packages("dplyr")
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 3.4.2
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.4.2
```

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

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

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

```
library(ggplot2)
library(sqldf)
```

```
## Loading required package: gsubfn
```

```
## Loading required package: proto
```

```
## Warning in doTryCatch(return(expr), name, parentenv, handler): unable to load shared
object '/Library/Frameworks/R.framework/Resources/modules//R_X11.so':
##   dlopen(/Library/Frameworks/R.framework/Resources/modules//R_X11.so, 6): Library not
loaded: /opt/X11/lib/libfontconfig.1.dylib
##   Referenced from: /Library/Frameworks/R.framework/Resources/modules//R_X11.so
##   Reason: Incompatible library version: R_X11.so requires version 11.0.0 or later, bu
t libfontconfig.1.dylib provides version 10.0.0
```

```
## Could not load tcltk.  Will use slower R code instead.
```

```
## Loading required package: RSQLite
```

```
library(magrittr)
```

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

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

Silver and Gold price data are loaded from the datasets

```
gold_df <- read.csv(file="/Users/anjalhussan/Desktop/data_science/proposal/GOLD.csv", he
ad=TRUE, sep=",",stringsAsFactors = FALSE)
head(gold_df)
```

```
##         Date USD..AM. USD..PM. GBP..AM. GBP..PM. EURO..AM. EURO..PM.
## 1 2017-11-30  1282.15  1280.20   952.64   948.88   1084.06   1074.98
## 2 2017-11-29  1294.85  1283.85   965.70   957.50   1092.46   1085.11
## 3 2017-11-28  1293.90  1291.85   972.75   974.18   1088.95   1087.61
## 4 2017-11-27  1294.70  1294.90   969.73   969.36   1084.83   1085.00
## 5 2017-11-24  1289.15  1290.50   967.89   966.58   1086.37   1082.60
## 6 2017-11-23  1290.15  1290.35   969.93   969.96   1089.40   1089.18
```

```
silver_df = read.csv(file="/Users/anjalhussan/Desktop/data_science/proposal/SILVER.csv",
 head=TRUE, sep=",",stringsAsFactors = FALSE)
head(silver_df)
```

```
##         Date    USD   GBP  EURO
## 1 2017-11-30 16.570 12.32 14.00
## 2 2017-11-29 16.895 12.60 14.26
## 3 2017-11-28 17.070 12.84 14.36
## 4 2017-11-27 17.100 12.81 14.32
## 5 2017-11-24 17.050 12.80 14.38
## 6 2017-11-23 17.095 12.84 14.43
```

Data Transformation, Cleanup and Preparation:

Apply tidyr select function and select only columns (Date and USD..AM Price) form gold_df data set that are relevant for our purpose

```
select(gold_df)
```

```
## data frame with 0 columns and 12618 rows
```

```
gold_dataSet = select (gold_df, matches("Date|USD..AM"))
```

Apply tidyr mutate functin to identify the price of the metal as gold

```
gold_dataSet = mutate (gold_dataSet, MetalPrice = "GoldPrice")
colnames(gold_dataSet)[2] <- "USD"
head(gold_dataSet)
```

```
##         Date     USD MetalPrice
## 1 2017-11-30 1282.15  GoldPrice
## 2 2017-11-29 1294.85  GoldPrice
## 3 2017-11-28 1293.90  GoldPrice
## 4 2017-11-27 1294.70  GoldPrice
## 5 2017-11-24 1289.15  GoldPrice
## 6 2017-11-23 1290.15  GoldPrice
```

Apply tidyr select function and select only columns (Date and USD Price) form silver_df data set that are relevant for our purpose. Also Apply tidyr mutate function to identify the price of the metal as Silver

```
select(silver_df)
```

```
## data frame with 0 columns and 12628 rows
```

```
Silver_dataSet = select (silver_df, matches("Date|USD"))
silver_df$Date = as.Date(as.character(silver_df$Date))
Silver_dataSet = mutate (Silver_dataSet, MetalPrice = "SilverPrice")
head(Silver_dataSet)
```

```
##         Date    USD  MetalPrice
## 1 2017-11-30 16.570 SilverPrice
## 2 2017-11-29 16.895 SilverPrice
## 3 2017-11-28 17.070 SilverPrice
## 4 2017-11-27 17.100 SilverPrice
## 5 2017-11-24 17.050 SilverPrice
## 6 2017-11-23 17.095 SilverPrice
```

Now combine both data frames (Silver and Gold price data frames) into one consolidated data frame for analysis

```
Silver_dataSet <- Silver_dataSet[1:nrow(gold_dataSet),1:3]
gold_dataSet <- gold_dataSet[,1:3]
#do.call(rbind, gold_dataSet)
combinedDataSets = rbind(gold_dataSet, Silver_dataSet)
head(combinedDataSets)
```

```
##         Date     USD MetalPrice
## 1 2017-11-30 1282.15  GoldPrice
## 2 2017-11-29 1294.85  GoldPrice
## 3 2017-11-28 1293.90  GoldPrice
## 4 2017-11-27 1294.70  GoldPrice
## 5 2017-11-24 1289.15  GoldPrice
## 6 2017-11-23 1290.15  GoldPrice
```

Apply arrange function to sort the data by Date

```
combinedDataSets = arrange(combinedDataSets, Date)
head(combinedDataSets)
```

```
##         Date   USD MetalPrice
## 1 1968-01-02 35.18  GoldPrice
## 2 1968-01-03 35.16  GoldPrice
## 3 1968-01-04 35.14  GoldPrice
## 4 1968-01-05 35.14  GoldPrice
## 5 1968-01-08 35.14  GoldPrice
## 6 1968-01-09 35.14  GoldPrice
```

Apply spread function to spread the data for each metal by adding one column for Gold and another for Silver

```
spreadedTotal = spread(combinedDataSets, MetalPrice, USD)
head(spreadedTotal, 20)
```

```
##          Date GoldPrice SilverPrice
## 1  1968-01-02     35.18          NA
## 2  1968-01-03     35.16          NA
## 3  1968-01-04     35.14          NA
## 4  1968-01-05     35.14          NA
## 5  1968-01-08     35.14          NA
## 6  1968-01-09     35.14          NA
## 7  1968-01-10     35.15          NA
## 8  1968-01-11     35.17          NA
## 9  1968-01-12     35.18          NA
## 10 1968-01-15     35.18          NA
## 11 1968-01-16     35.19       2.069
## 12 1968-01-17     35.20       2.067
## 13 1968-01-18     35.20       2.058
## 14 1968-01-19     35.19       2.078
## 15 1968-01-22     35.19       2.088
## 16 1968-01-23     35.19       2.080
## 17 1968-01-24     35.20       2.064
## 18 1968-01-25     35.20       2.018
## 19 1968-01-26     35.20       2.069
## 20 1968-01-29     35.19       2.039
```

Create a function to calculate the price ratio and round it to 4 decimal digits

```
calcPriceRatio = function(goldPrice, silverPrice)
{
  round((goldPrice/silverPrice), 4)
  }
```

Apply tidyr mutate function to add PriceRatio and transactionYear columns

```
dataSetsWithPriceRatio = mutate(spreadedTotal, PriceRatio=calcPriceRatio(GoldPrice, Silv
erPrice), transactionYear=substring(Date,1,4))
tail(dataSetsWithPriceRatio, 20)
```

```
##             Date GoldPrice SilverPrice PriceRatio transactionYear
## 12611 2017-11-03   1275.30      17.085    74.6444            2017
## 12612 2017-11-06   1271.60      16.915    75.1759            2017
## 12613 2017-11-07   1276.35      17.005    75.0573            2017
## 12614 2017-11-08   1282.25      16.995    75.4487            2017
## 12615 2017-11-09   1284.00      17.100    75.0877            2017
## 12616 2017-11-10   1284.45      17.000    75.5559            2017
## 12617 2017-11-13   1278.40      16.925    75.5332            2017
## 12618 2017-11-14   1273.70      16.935    75.2111            2017
## 12619 2017-11-15   1285.70      17.115    75.1212            2017
## 12620 2017-11-16   1277.70      17.040    74.9824            2017
## 12621 2017-11-17   1283.85      17.085    75.1449            2017
## 12622 2017-11-20   1292.35      17.145    75.3777            2017
## 12623 2017-11-21   1280.00      17.000    75.2941            2017
## 12624 2017-11-22   1283.95      16.965    75.6823            2017
## 12625 2017-11-23   1290.15      17.095    75.4694            2017
## 12626 2017-11-24   1289.15      17.050    75.6100            2017
## 12627 2017-11-27   1294.70      17.100    75.7135            2017
## 12628 2017-11-28   1293.90      17.070    75.7996            2017
## 12629 2017-11-29   1294.85      16.895    76.6410            2017
## 12630 2017-11-30   1282.15      16.570    77.3778            2017
```

Select price data for years from 2001 to 2013

```
PriceDataYear_2001_2013 = filter(dataSetsWithPriceRatio, transactionYear %in% c("2001",
"2002", "2003", "2004", "2005", "2006", "2007", "2008", "2009", "2010", "2011", "2012",
"2013"))
tail(PriceDataYear_2001_2013, 20)
```

```
##            Date GoldPrice SilverPrice PriceRatio transactionYear
## 3266 2013-12-02   1237.50       19.75    62.6582            2013
## 3267 2013-12-03   1219.00       19.17    63.5889            2013
## 3268 2013-12-04   1213.00       19.05    63.6745            2013
## 3269 2013-12-05   1234.00       19.46    63.4121            2013
## 3270 2013-12-06   1230.75       19.49    63.1478            2013
## 3271 2013-12-09   1228.50       19.50    63.0000            2013
## 3272 2013-12-10   1245.75       20.05    62.1322            2013
## 3273 2013-12-11   1255.25       20.39    61.5620            2013
## 3274 2013-12-12   1243.50       19.80    62.8030            2013
## 3275 2013-12-13   1222.75       19.55    62.5448            2013
## 3276 2013-12-16   1229.50       19.50    63.0513            2013
## 3277 2013-12-17   1237.25       20.02    61.8007            2013
## 3278 2013-12-18   1233.25       19.94    61.8480            2013
## 3279 2013-12-19   1205.25       19.34    62.3190            2013
## 3280 2013-12-20   1195.00       19.33    61.8210            2013
## 3281 2013-12-23   1192.75       19.37    61.5772            2013
## 3282 2013-12-24   1196.50       19.40    61.6753            2013
## 3283 2013-12-27   1209.25       19.92    60.7053            2013
## 3284 2013-12-30   1201.50       19.65    61.1450            2013
## 3285 2013-12-31   1201.50       19.50    61.6154            2013
```

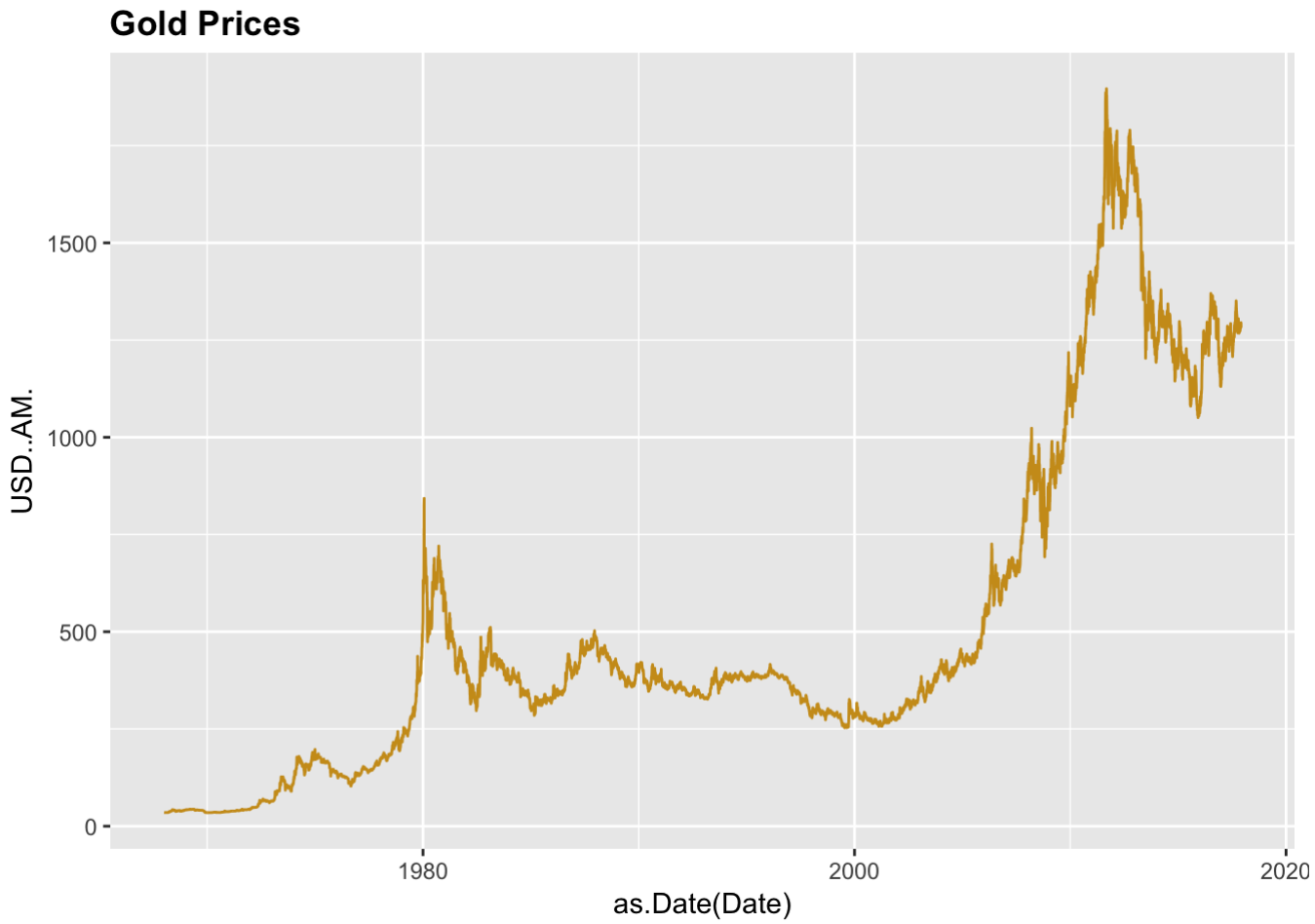
```
summary(PriceDataYear_2001_2013)
```

```
##      Date              GoldPrice       SilverPrice       PriceRatio
##  Length:3285        Min.   : 256.7   Min.   : 4.065   Min.   :31.44
##  Class :character   1st Qu.: 399.3   1st Qu.: 6.200   1st Qu.:53.19
##  Mode  :character   Median : 673.6   Median :12.850   Median :60.22
##                     Mean   : 832.2   Mean   :14.807   Mean   :60.04
##                     3rd Qu.:1243.0   3rd Qu.:19.650   3rd Qu.:65.73
##                     Max.   :1896.5   Max.   :48.700   Max.   :83.79
##  transactionYear
##  Length:3285
##  Class :character
##  Mode  :character
##
##
##
```

Data Visualization:

Gold Price Data

```
ggplot (gold_df, aes(as.Date(Date), USD..AM.)) +
  geom_line (aes(color="Gold")) +
  labs (color="Legend") +
  scale_colour_manual ("", breaks = c("gold"), values = c("goldenrod3")) +
  ggtitle ("Gold Prices") + theme (plot.title = element_text(lineheight=0.7, face="bold"
))
```

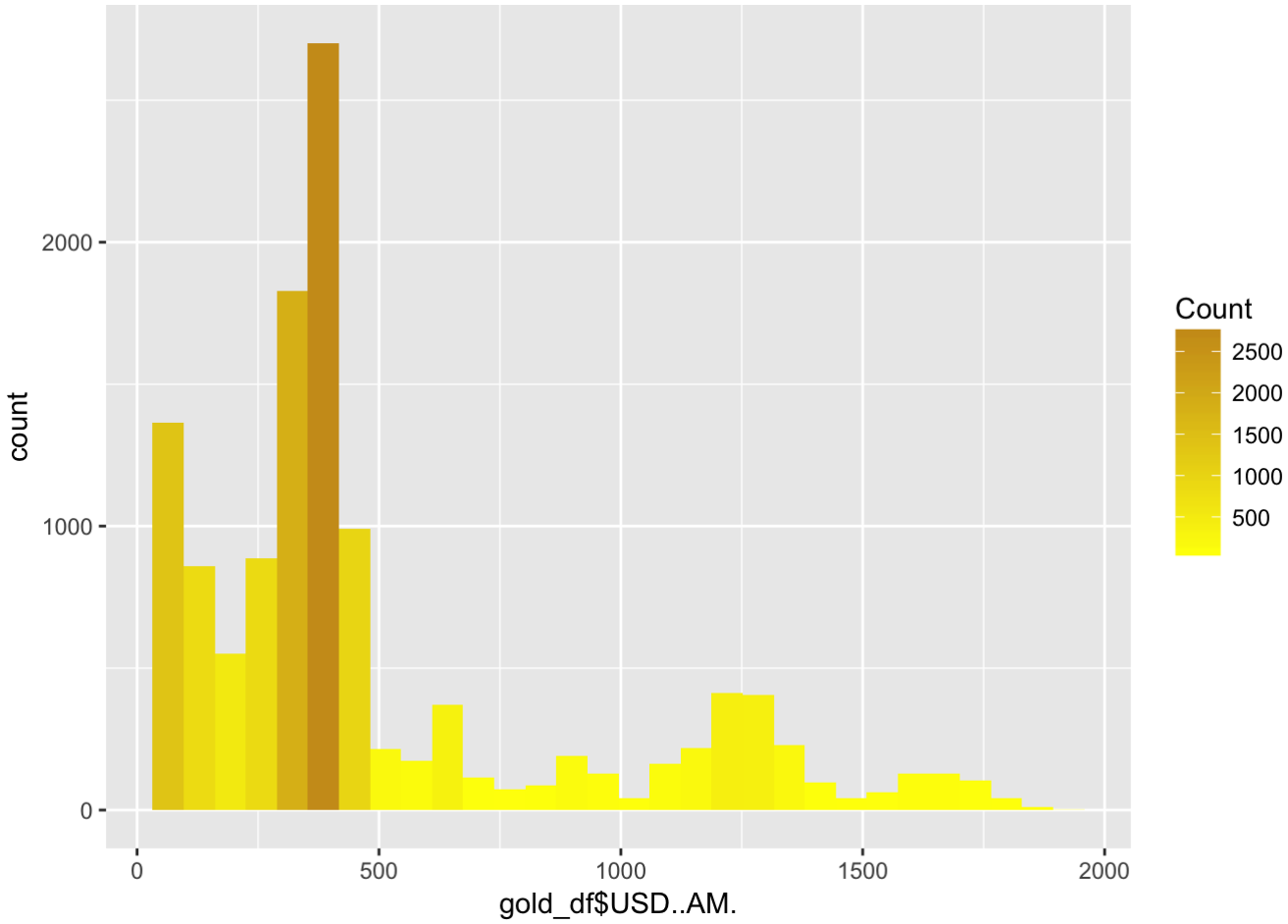


Gold Price Histogram

```
ggplot(data=gold_df, aes(gold_df$USD..AM.)) +
  geom_histogram(aes(fill = ..count..)) +
  scale_fill_gradient("Count", low = "yellow", high = "goldenrod3")
```

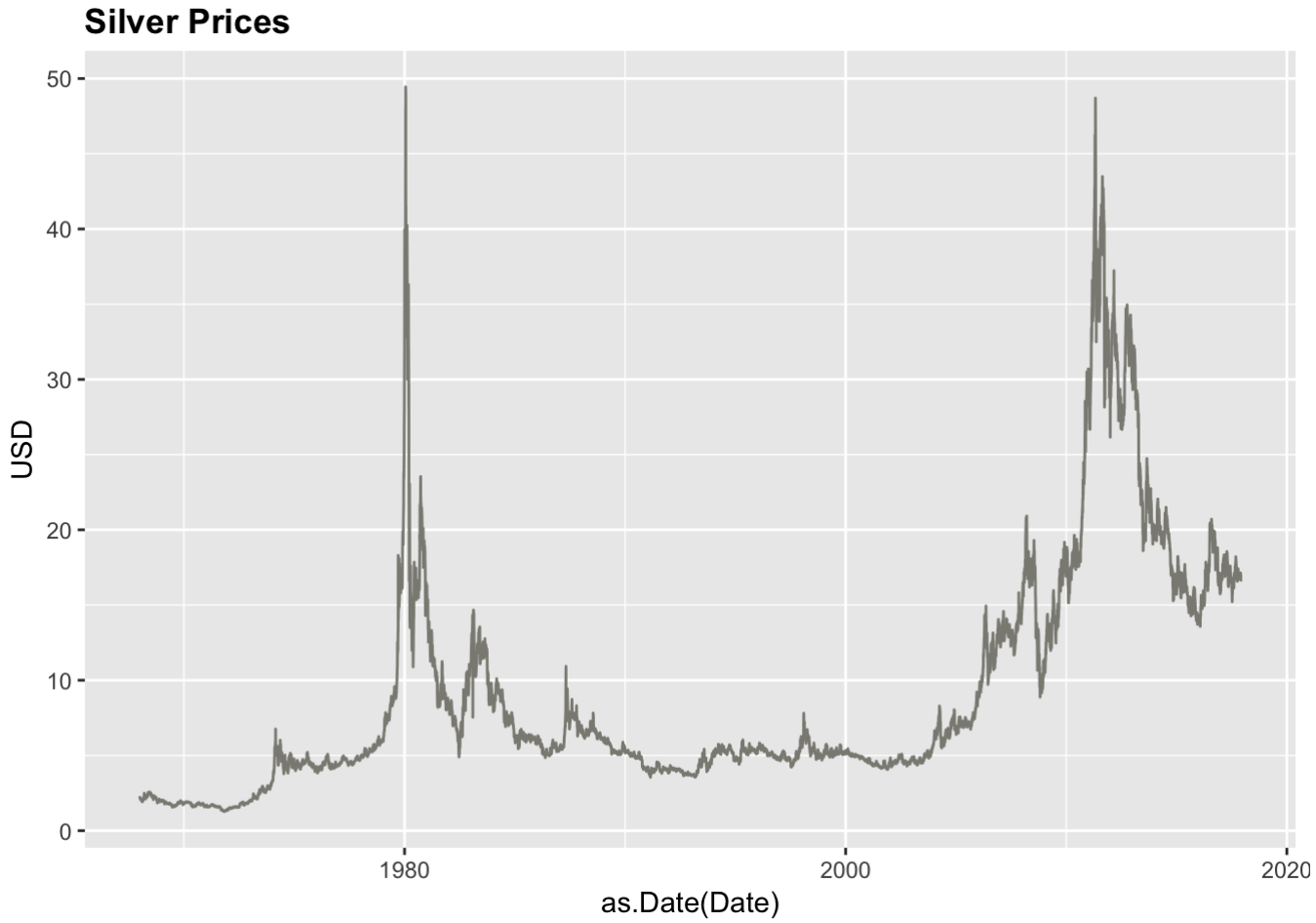
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```



Silver Price Data

```
ggplot (silver_df, aes(as.Date(Date), USD)) +
  geom_line (aes(color="Silver")) +
  labs (color="Legend") +
  scale_colour_manual ("", breaks = c("silver"), values = c("ivory4")) +
  ggtitle ("Silver Prices") +
  theme (plot.title = element_text(lineheight=0.7, face="bold"))
```

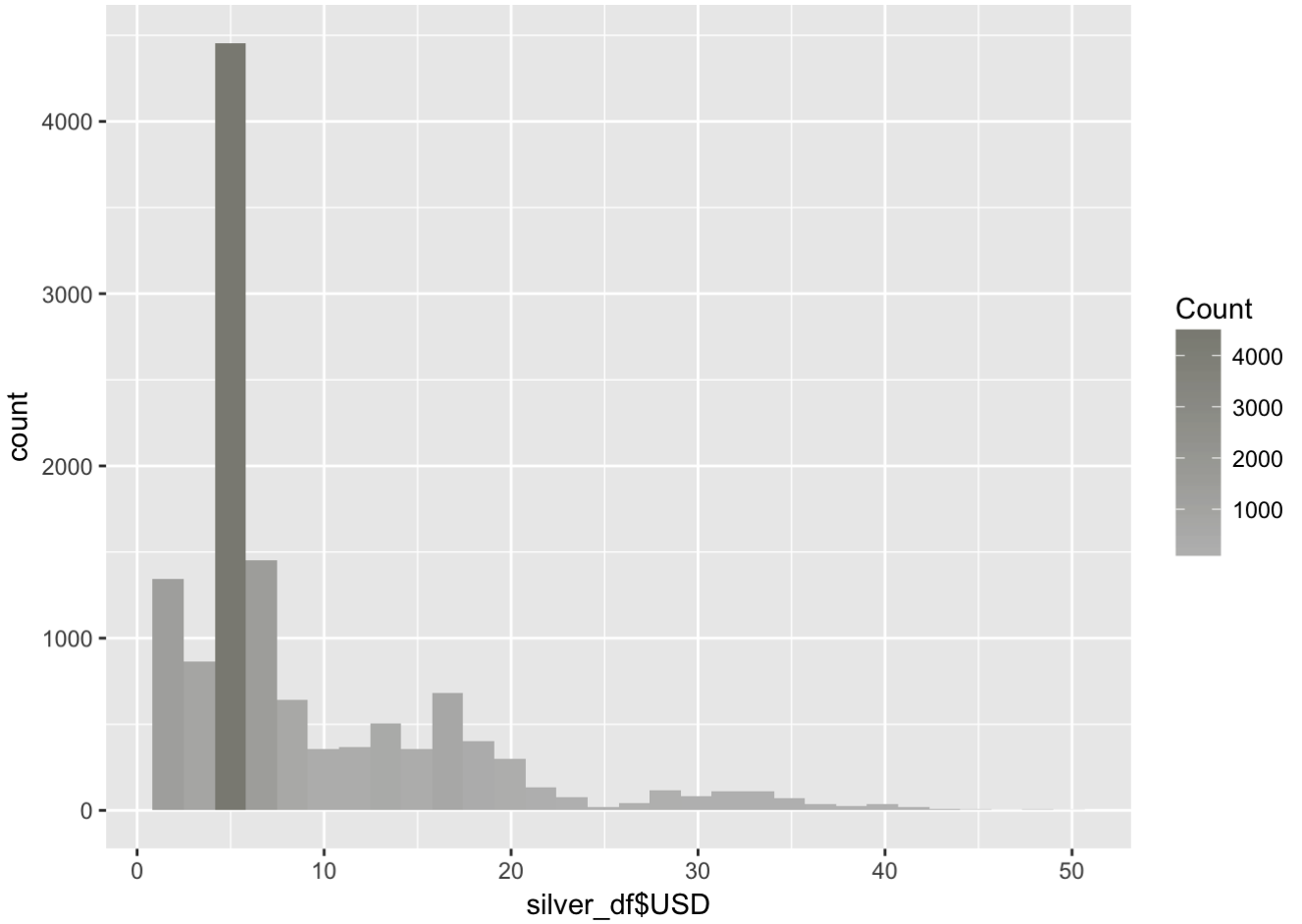


Silver Price Histogram

```
ggplot(data=silver_df, aes(silver_df$USD)) +
  geom_histogram(aes(fill = ..count..)) +
  scale_fill_gradient("Count", low = "grey", high = "ivory4")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 19 rows containing non-finite values (stat_bin).
```



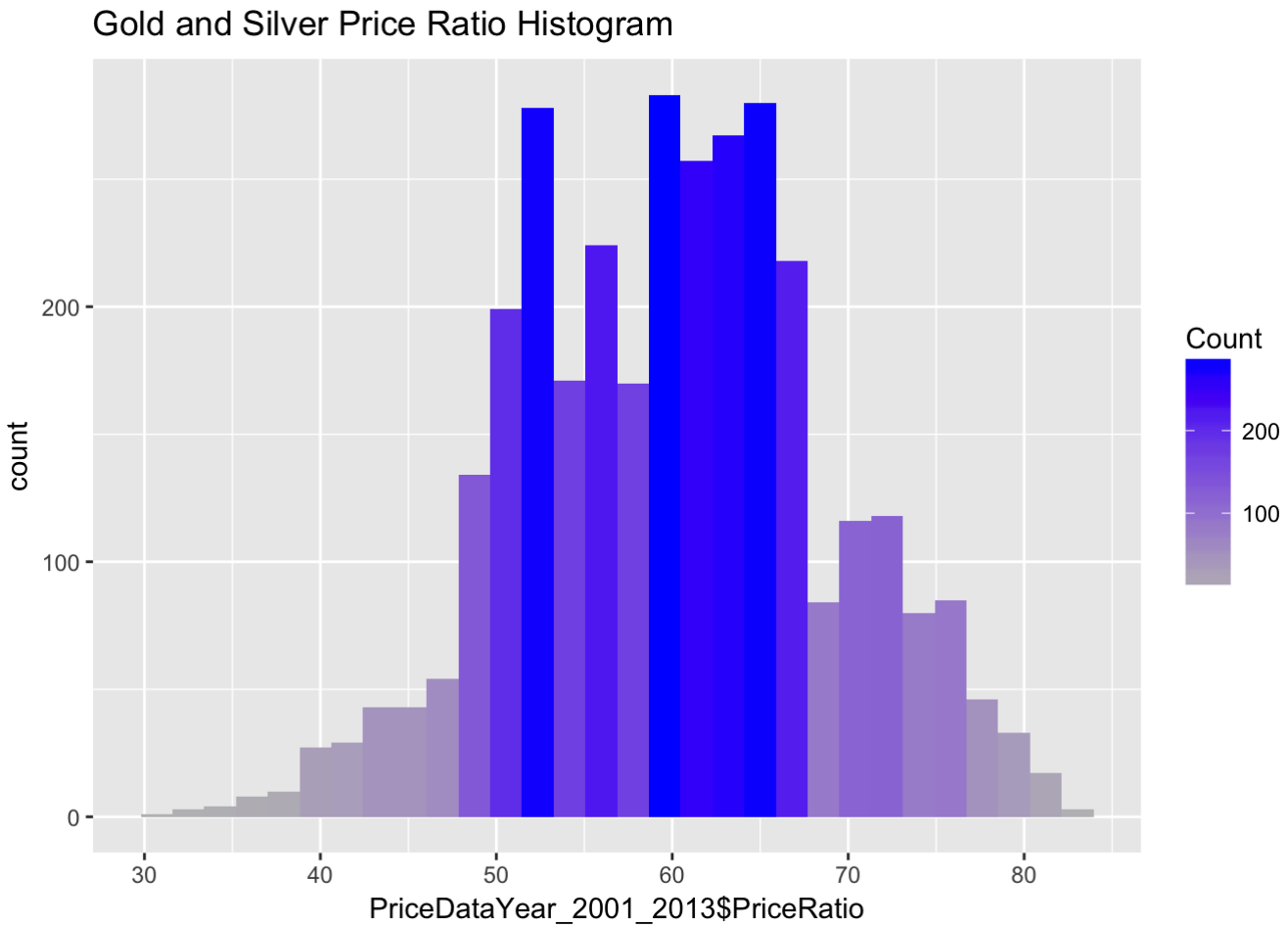
Gold/Silver Price Ratio Data

```
ggplot (PriceDataYear_2001_2013, aes(as.Date(Date), PriceRatio)) +
  geom_line (aes(color="Blue")) +
  labs (color="Legend") +
  scale_colour_manual ("", breaks = c("gold"), values = c("goldenrod3")) +
  ggtitle ("Gold and Silver Price Ratio") +
  theme (plot.title = element_text(lineheight=0.7, face="bold"))
```

Gold/Silver Price Ratio Histogram

```
ggplot(data=PriceDataYear_2001_2013, aes(PriceDataYear_2001_2013$PriceRatio)) +
  geom_histogram(aes(fill = ..count..)) +
  scale_fill_gradient("Count", low = "grey", high = "blue") +
  ggtitle ("Gold and Silver Price Ratio Histogram")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Statistical Analysis:

In this section we will create a linear regression model and calculate the correlation between the data to see if there is a strong relationship between silver and gold prices.

Create a function to calculate the correlation and round it to 4 decimal digits

```
findCorrelation <- function() {
  x = PriceDataYear_2001_2013$SilverPrice
  y = PriceDataYear_2001_2013$GoldPrice
  corr = round(cor(x, y),4)
  print (paste0("Correlation = ",corr))
  return (corr)
}

c = findCorrelation()
```

```
## [1] "Correlation = 0.9643"
```

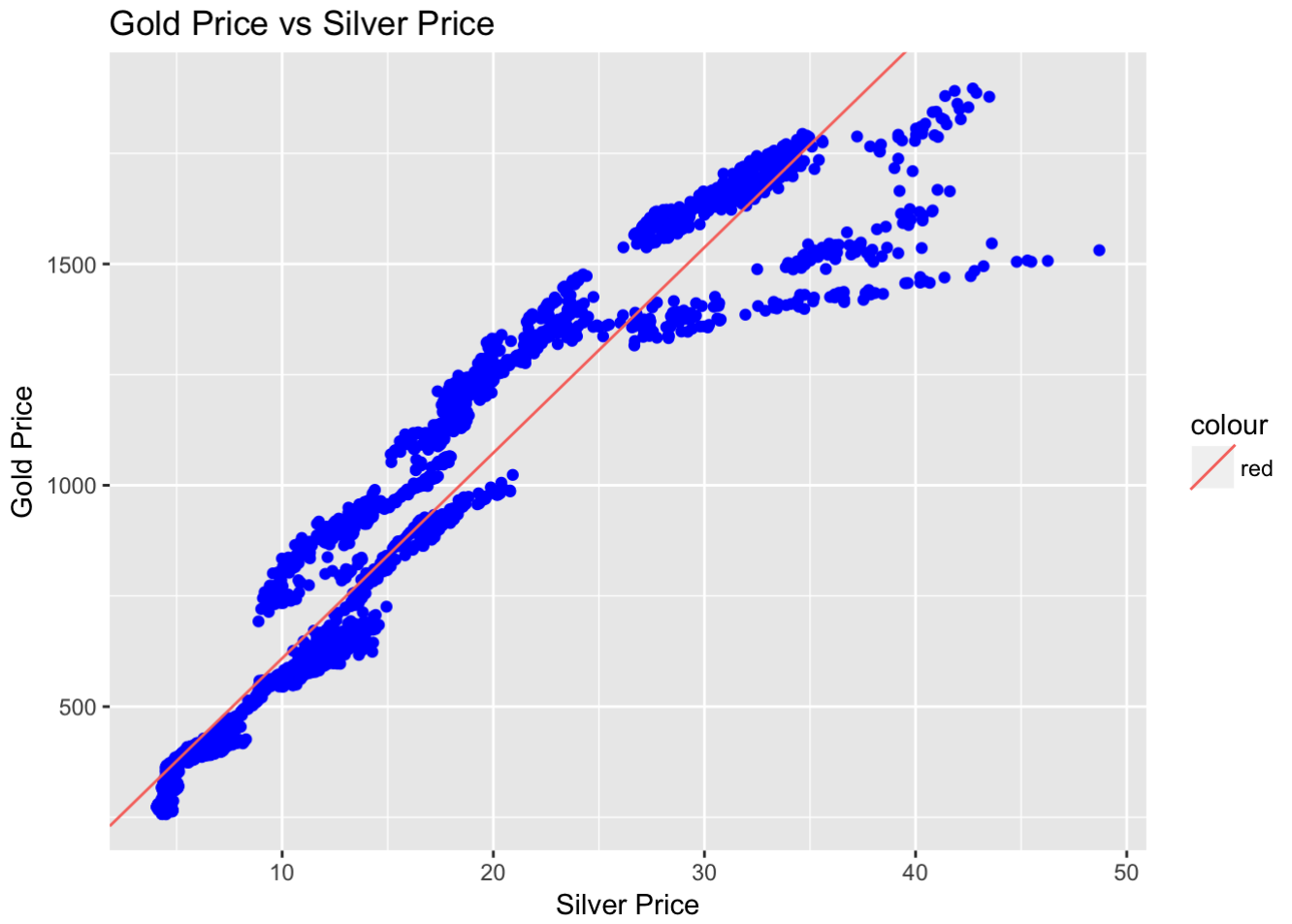
Create a function for Linear Model

```
findStatsFunction <- function() {
  m = lm (GoldPrice ~ SilverPrice, data = PriceDataYear_2001_2013)
  s = summary(m)
  print(s)

  slp = round(m$coefficients[2], 4)
  int = round(m$coefficients[1], 4)

  return (m)
}
m = findStatsFunction()
```

```
##
## Call:
## lm(formula = GoldPrice ~ SilverPrice, data = PriceDataYear_2001_2013)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -874.17  -63.04  -28.61  100.34  263.92
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 144.9725     3.9801   36.42   <2e-16 ***
## SilverPrice  46.4107     0.2224  208.64   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 128.1 on 3283 degrees of freedom
## Multiple R-squared:  0.9299, Adjusted R-squared:  0.9298
## F-statistic: 4.353e+04 on 1 and 3283 DF,  p-value: < 2.2e-16
```

Display the Linear Model

```
plot = ggplot(PriceDataYear_2001_2013, aes(SilverPrice, GoldPrice)) + geom_point(colour=
"blue") +
    xlab("Silver Price") + ylab("Gold Price") + labs(title = "Gold Price vs Silver Pric
e") +
    geom_abline(aes(slope=round(m$coefficients[2], 4), intercept=round(m$coefficients[1]
, 4), color="red"))
  print(plot)
```



Regression Statistics

Linear Regression Equation: goldPrice = 144.9725 + (46.4107 * silverPrice) Correlation Coefficient: 0.9643 Multiple R-Square: 0.9299 Adjusted R-Square: 0.9298 Description: Strong correlation. Model fits the data

Hypothesis Testing

H_0 : Null Hypothesis: There is no relationship between silver and gold prices H_A : Alternative Hypothesis: There is a relationship between silver and gold prices

The multiple R value is 0.9299 which indicats that there is significant correlation between silver and gold prices. The value of Adjusted R square is 0.9298 which also indicates that silver price affects the gold price. Therefore, we reject the null hypothesis (H_0) and accept the Alternative hypothesis (H_1).

Conclusion:

The two variables (silver price and gold Price) change in the same direction. If the silver price increases the gold price increases as well. Therefore, there is a positive correlation of 0.9641 between the two variables. Also, from the linear regression model, we can reject the null hypothesis and accept the alternative hypothesis. In conclusion, there is a strong relationship between silver and gold prices for the 13 year period of study (2001-2013).

End of this project file