How many bows are there compared to total sales?

We are working with the Tarisio data set, which includes a total of 52,250 entries. This data set includes the sales of different violin types, makers, etc. We are interested in violin bow sales, which account for 16,134 of the recorded sales.

How many nulls/zeroes?

The data is not perfect, for some entries there are missing pieces, such as sale price, maker, auction house, etc. The following table shows all of the null counts for the non-numerical columns:

Category	***Date***	***Latest Date***	***Info***
Null Count #	1523	1523	750
Category	***USD***	***EUR***	***GBP***
Zero Count #	1	6025	1

It is my understanding that the ‘Date’ or ‘Latest Date’ have the same number of nulls, due to the fact that the ‘Latest Date’ column pulls the most recent date from the ‘Date’ column. So when there is no date recorded, the ‘Latest Date’ column is also null.

For the zeroes, we can use USD or GBP to avoid a lack of data, and we can choose to either remove the one zero, or see if we can convert one of the currencies to USD or GBP if that information is available.

What currency looks the best to use?

USD looks best to use because the values are in place and have very few null values.

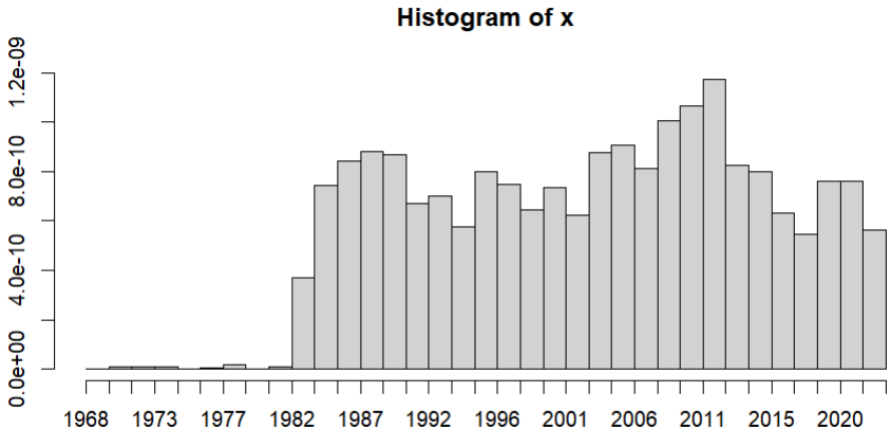
Which makers or non numerical values occur a lot?

Maker	**Total**
Hill & Sons, W. E. Firm	771
Sartory, Eug√®ne Nicolas	372
Tubbs, James	276
Voirin, Fran√ßois Nicolas	171
Pfretzschner, Hermann Richard	147

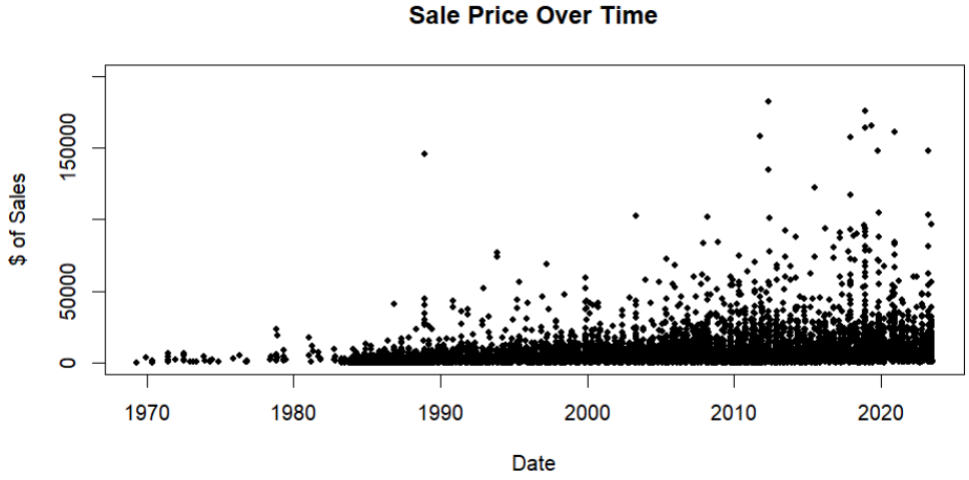
Thomassin, Claude Auguste	120
N√ºrnberger, Albert	114
Lamy, Joseph Alfred p√®re	110
F√©tique, Victor Fran√ßois	109

Minimum/Maximum for Date and Sales Price, along with summary of results (mean, standard dev, etc) -

The date range for this set of data is ‘1969-04-24’ to ‘2023-07-07’, which is a range of just about 54 years. It appears that right around 1982, the amount of recorded sales rose dramatically:



The sales price also appears to be slowly increasing over time:



After running a summary in R on both the sale date and sale price, the following was gathered:

	Minimum	Median	Maximum	Mean	1st Qu.	3rd Qu.
Sale Date	1969-04-24	2004-10-17	2023-07-07	2003-07-13	1993-04-29	2012-04-27
Sale Price (USD)	0	2400	367862	5147	1176	5018

When looking at the sale dates, the histogram paints a better picture I believe. We can see that the density of sales increases around 1982, and since we are looking for long-term investments, we should include as much data as possible, so starting at this 'spike' may be ideal.

Add Into the Code

How does removing the zeroes affect the sales results? Should we remove these entirely? Show side by side. -

After removing the zeros from the data, only 1 entry is removed from the USD column. The following is a summary of the sale prices after removing the zero:

	Minimum	Median	Maximum	Mean	1st Qu.	3rd Qu.
Sale Price (USD)	24	3156	367862	6552	1560	7020

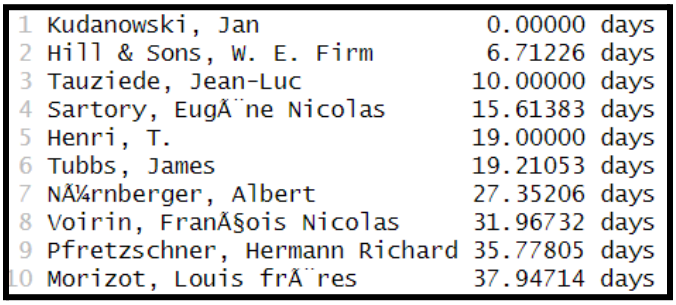
Removing a single zero did change the results, the median, mean, and quantiles all rose due to removing the zero. The changes seem quite significant in this case.

Find number of columns that are numerical and non numerical -

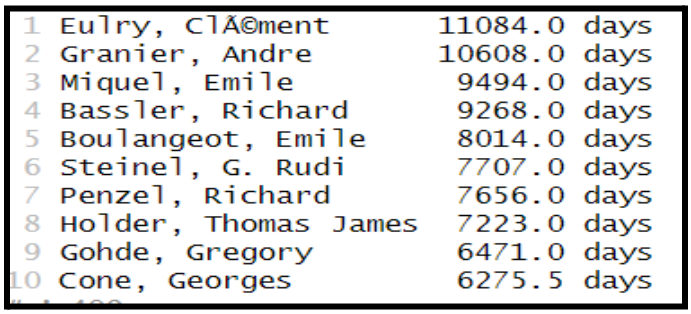
Numeric	**Non Numeric**
4	10

That is, we want to calculate the time between each sale per maker, and then take the average of those differences *per maker*. Then, you should rank it from lowest to

highest. Remember, it's good to know the average holding time between sale to sale per maker. - Ryan

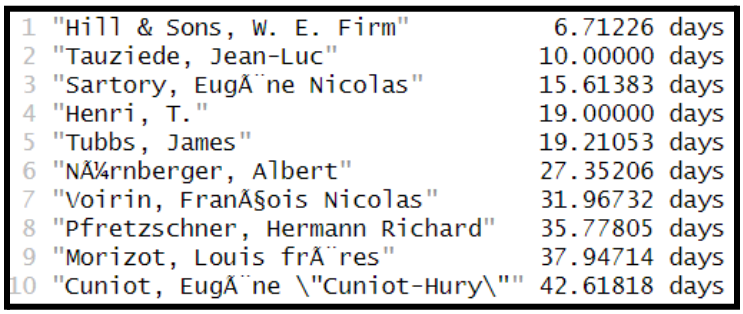


1	Kudanowski, Jan	0.00000 days
2	Hill & Sons, W. E. Firm	6.71226 days
3	Tauziede, Jean-Luc	10.00000 days
4	Sartory, EugÃ¨ne Nicolas	15.61383 days
5	Henri, T.	19.00000 days
6	Tubbs, James	19.21053 days
7	NÃ¼rnberger, Albert	27.35206 days
8	Voirin, FranÃ§ois Nicolas	31.96732 days
9	Pfretzschner, Hermann Richard	35.77805 days
10	Morizot, Louis frÃ¨res	37.94714 days



1	Eulry, ClÃ©ment	11084.0 days
2	Granier, Andre	10608.0 days
3	Miquel, Emile	9494.0 days
4	Bassler, Richard	9268.0 days
5	Boulangeot, Emile	8014.0 days
6	Steinel, G. Rudi	7707.0 days
7	Penzel, Richard	7656.0 days
8	Holder, Thomas James	7223.0 days
9	Gohde, Gregory	6471.0 days
10	Cone, Georges	6275.5 days

The average sale time per maker varies greatly in this data set. The higher end averages are anywhere from 6,275 days to 11,084 days between sales. For bows being sold at a higher rate, the averages are anywhere from 6.7 days to 37.9 days between sales.



1	"Hill & Sons, W. E. Firm"	6.71226 days
2	"Tauziede, Jean-Luc"	10.00000 days
3	"Sartory, EugÃ¨ne Nicolas"	15.61383 days
4	"Henri, T."	19.00000 days
5	"Tubbs, James"	19.21053 days
6	"NÃ¼rnberger, Albert"	27.35206 days
7	"Voirin, FranÃ§ois Nicolas"	31.96732 days
8	"Pfretzschner, Hermann Richard"	35.77805 days
9	"Morizot, Louis frÃ¨res"	37.94714 days
10	"Cuniot, EugÃ¨ne \"Cuniot-Hury\""	42.61818 days

After removing the zeroes, the new range of low averages is now from 6.7 days to 42.6 days between sales.

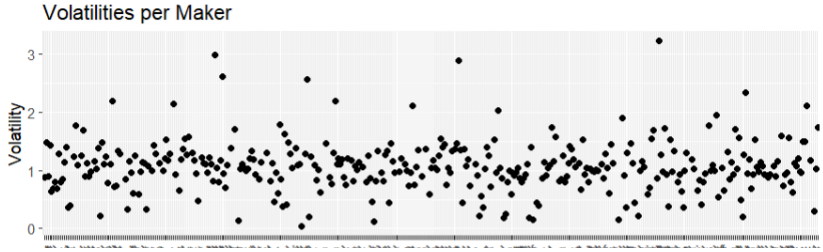
For Next Meeting:

- **Find relationship between log returns and holding period**
- **How volatile is the market for each maker? -**

1	Schubert, Paul	3.23
2	Dolmetsch, Arnold	2.98
3	Lyon & Healy, Firm	2.89
4	Duchaine, Nicolas II	2.62
5	Grimm, Karl	2.57
6	Tubbs, Edward	2.34
7	Hermann, Edwin Lothar	2.19
8	Bisch, Paul	2.19
9	Collins, Roy	2.14
10	Lamy, Alfred II	2.11

1	Grandchamp, Eric	0.0318
2	Key, Albert E.	0.121
3	Eury, Francois	0.135
4	Reichert, Eduard	0.155
5	Neuner, Ludwig	0.155
6	Nehr, Gilles	0.178
7	Morizot, AndrÃ© Auguste	0.179
8	Guillaume, Pierre	0.195
9	Tubbs, Family	0.203
10	Salchow, Isaac	0.211

- **Look at distributions, plot them in R -**

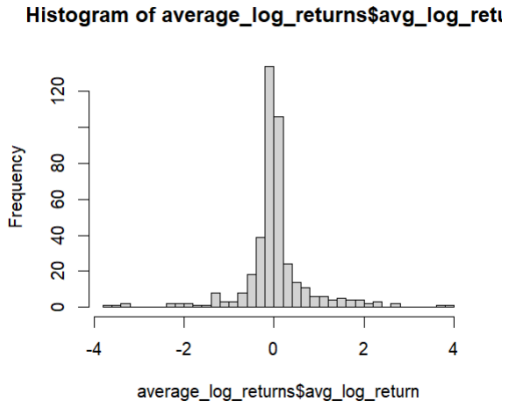


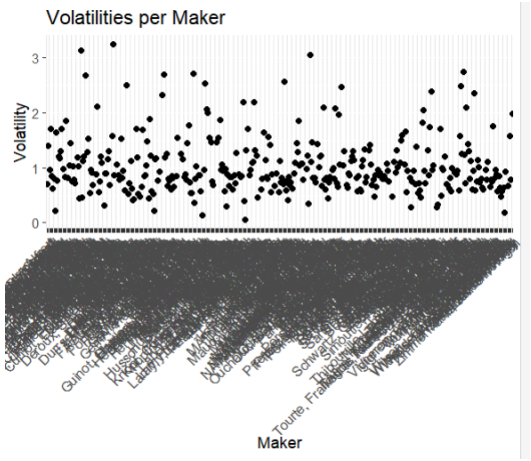
- **Find summary (mean, median, etc) for all bows to act as a benchmark -**

Mean	5983.75
Median	3450.00
Minimum	69.00
Max	147982.00

For our Next Meeting:

- scatterplot of returns and histogram of total returns -





-

- Find groupings of makers, see how that investment compares to benchmark overall. For example, if you were to invest in 3 specific makers, how would the return and volatility compare to the benchmark? (Select the top 3 number of makers)--

	RETURN	VOLATILITY
Hill & Sons, W. E. Firm	-0.0010075419	0.6754464
Sartory, EugÃ¨ne Nicolas	0.0004332577	0.94477542
Tubbs, James	-9.355680e-04	0.7950368
Benchmark	2.85%	0.523%

-

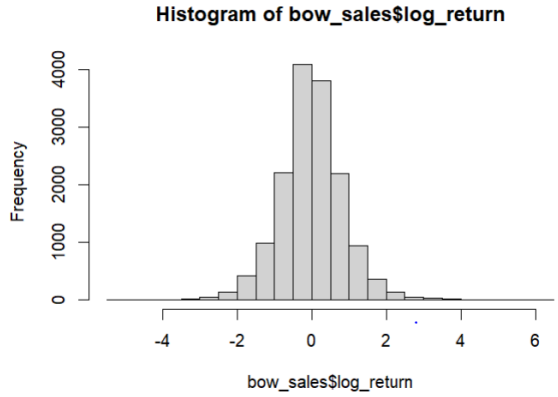
Next Meeting:

- Find number of sales for each of the makers, look at return data of makers with a lot of volume.-

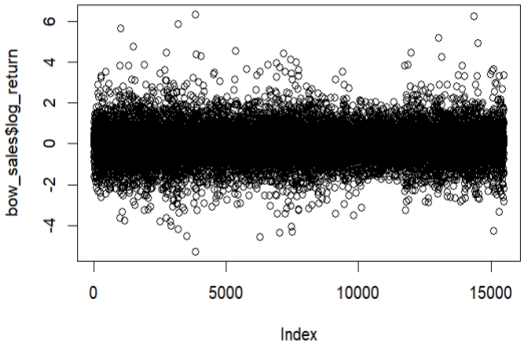
Name of the Maker	**Sales in USD**
F√©tique, Victor Fran√ßois	761968
Hill & Sons, W. E. Firm	2465577
Lamy, Joseph Alfred p√®re	947891
Peccatte, Dominique	1554362

Sartory, Eug√®ne Nicolas	4723056
Simon, Pierre	792848
Tourte, Fran√ßois Xavier 'le Jeune'	1015980
Tubbs, James	1689689
Voirin, Fran√ßois Nicolas	1439317

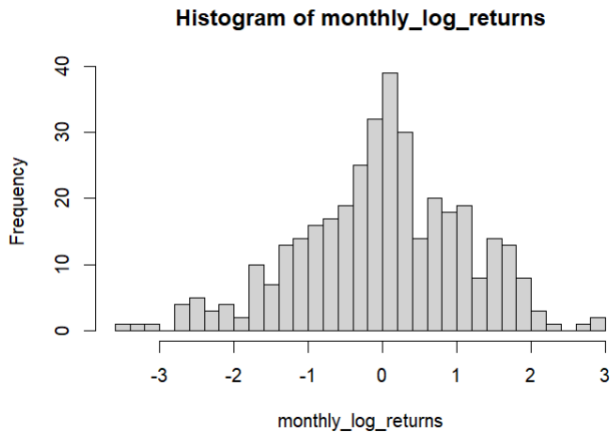
- Updated plot of of log returns, show all log returns, not just average per maker. -



-



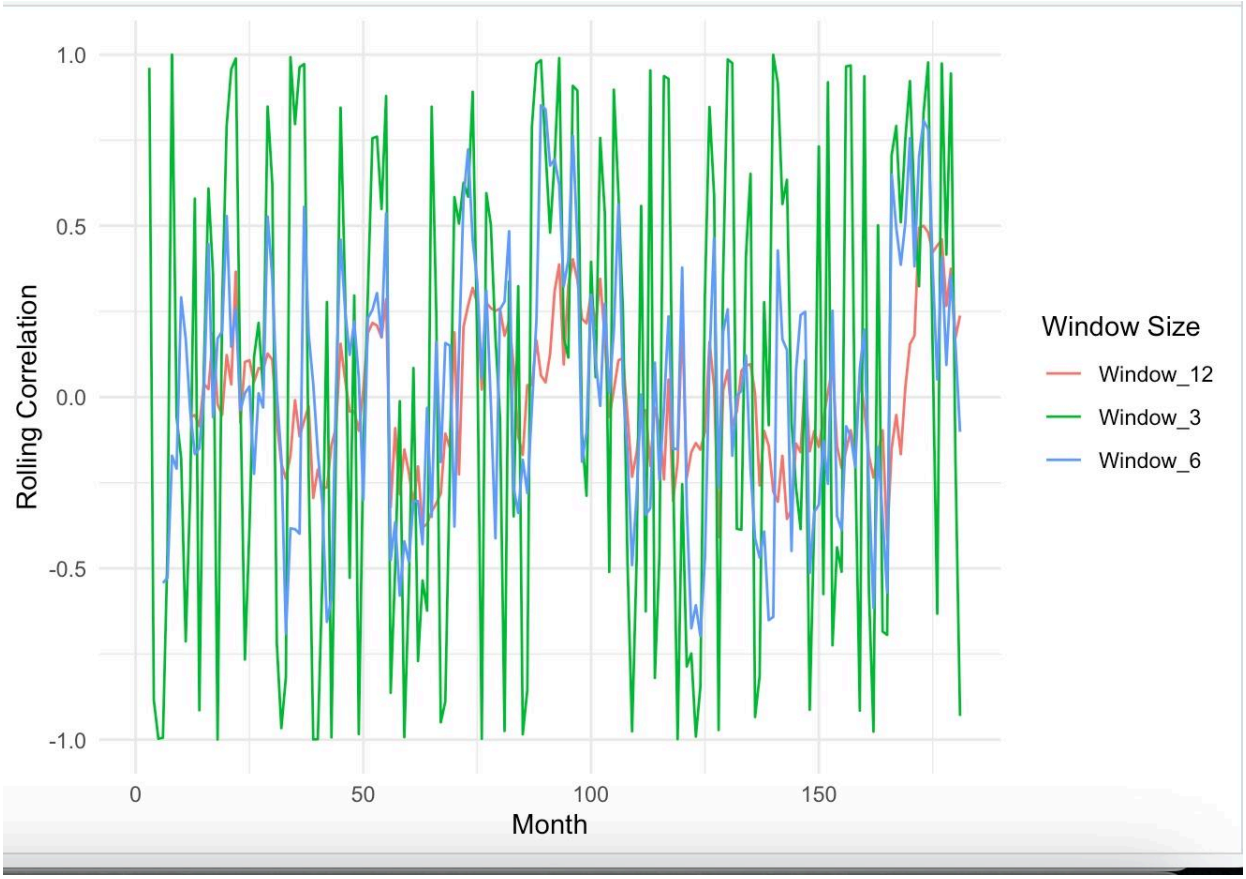
- Find benchmark in terms of total price in the market. Turn returns into monthly form, average price per month for all makers, log return over previous month -

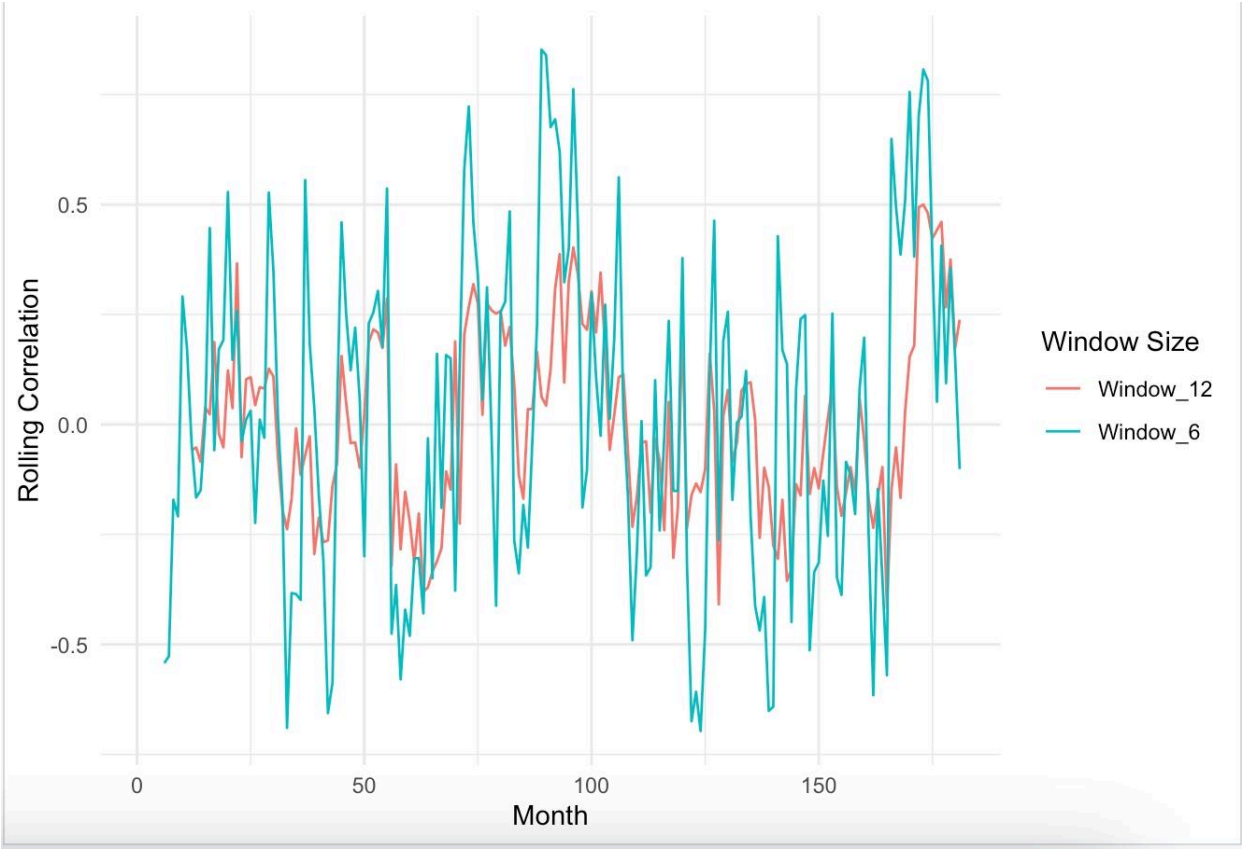


-
    - Average Monthly Log Return of Benchmark = -.0.004922416
    - Volatility/SD = 1.112424%
- Inflation data, turn into monthly, group by year and month and then take the average of those to find monthly average inflation data. -
    - We have 252 rows or monthly inflation data points to use.
- Start running correlations between inflation and the returns. Make sure both are monthly, print the results.-

```
> print(correlation)
                avg_T10YIE monthly_return
avg_T10YIE      1.00000000    -0.03504885
monthly_return -0.03504885     1.00000000
```

- Start working on the rolling correlations, picking a few look-back windows.(set the window for 3,6,12)--


1.0
0.5
0.0
-0.5
-1.0
Rolling Correlation
0
50
100
150
Month
Window Size
Window_12
Window_3
Window_6

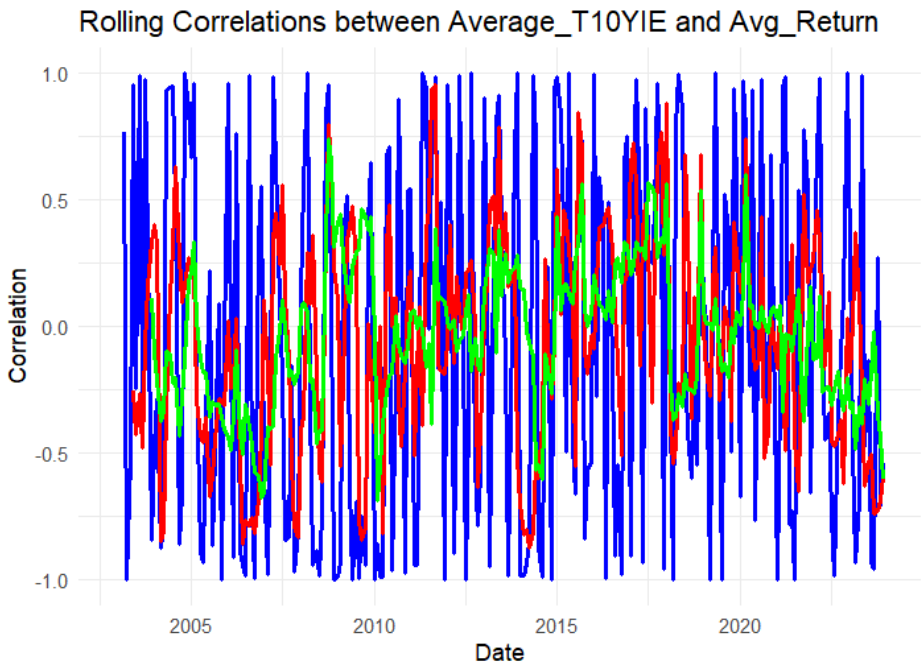


FOR NEXT MEETING 4/18

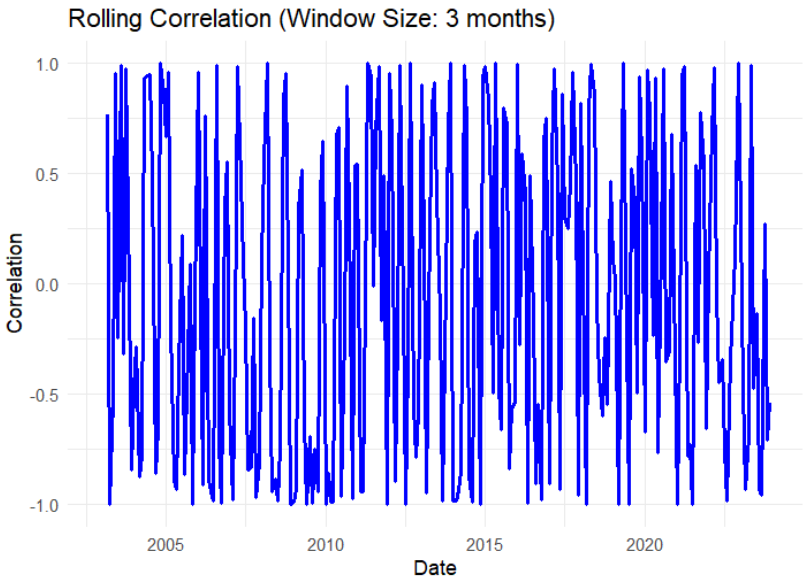
- Do the same rolling correlation windows with the T10 data and S&P Index Returns -

```
> correlation
[1] 0.05773641
```

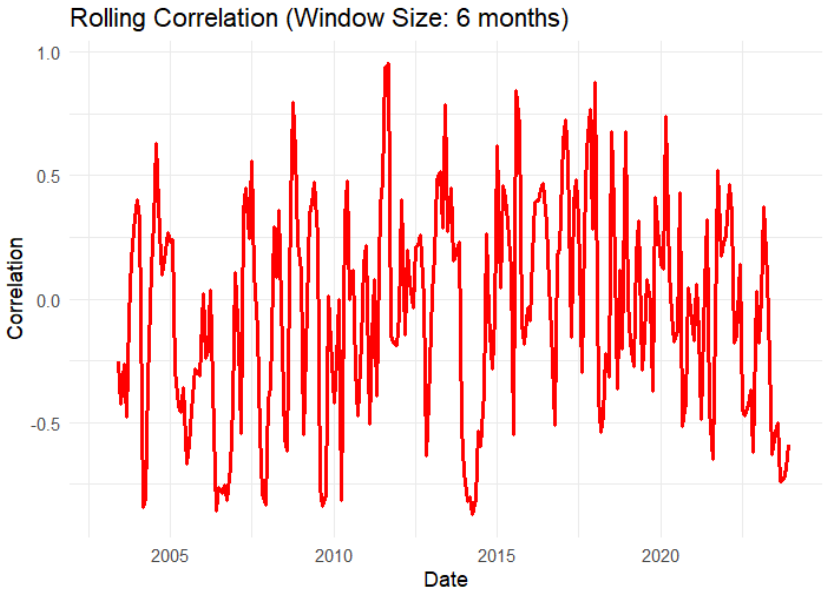
-

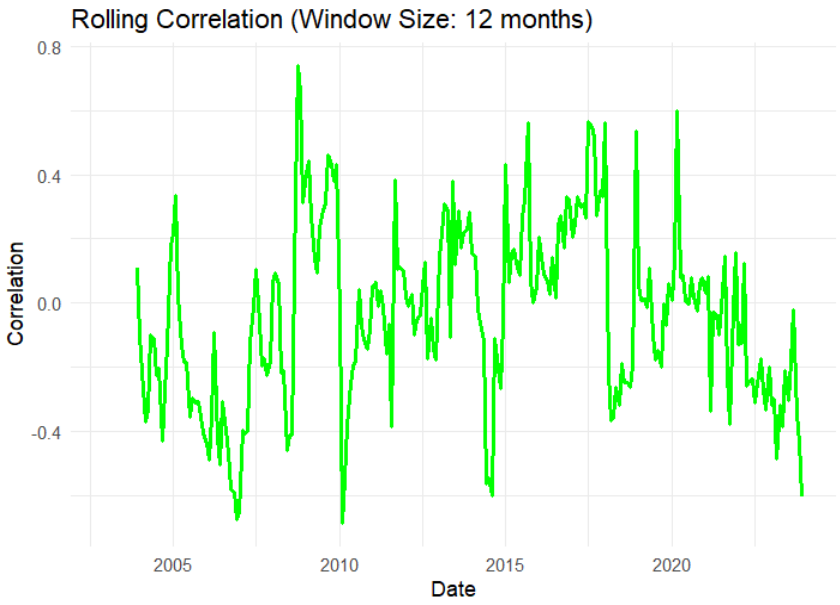

Rolling Correlations between Average_T10YIE and Avg_Return
1.0
0.5
0.0
-0.5
-1.0
Correlation
2005
2010
2015
2020
Date

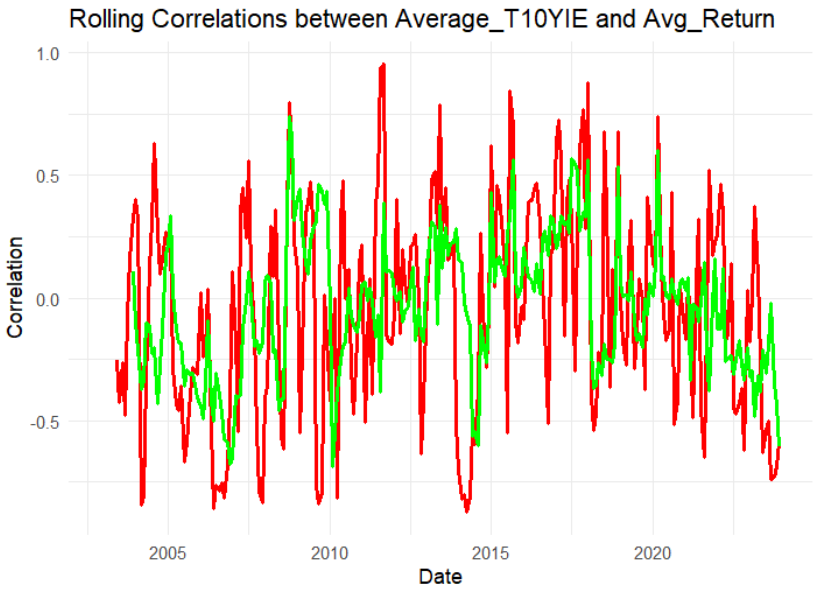
-


Rolling Correlation (Window Size: 3 months)
1.0
0.5
0.0
-0.5
-1.0
Correlation
2005
2010
2015
2020
Date

-

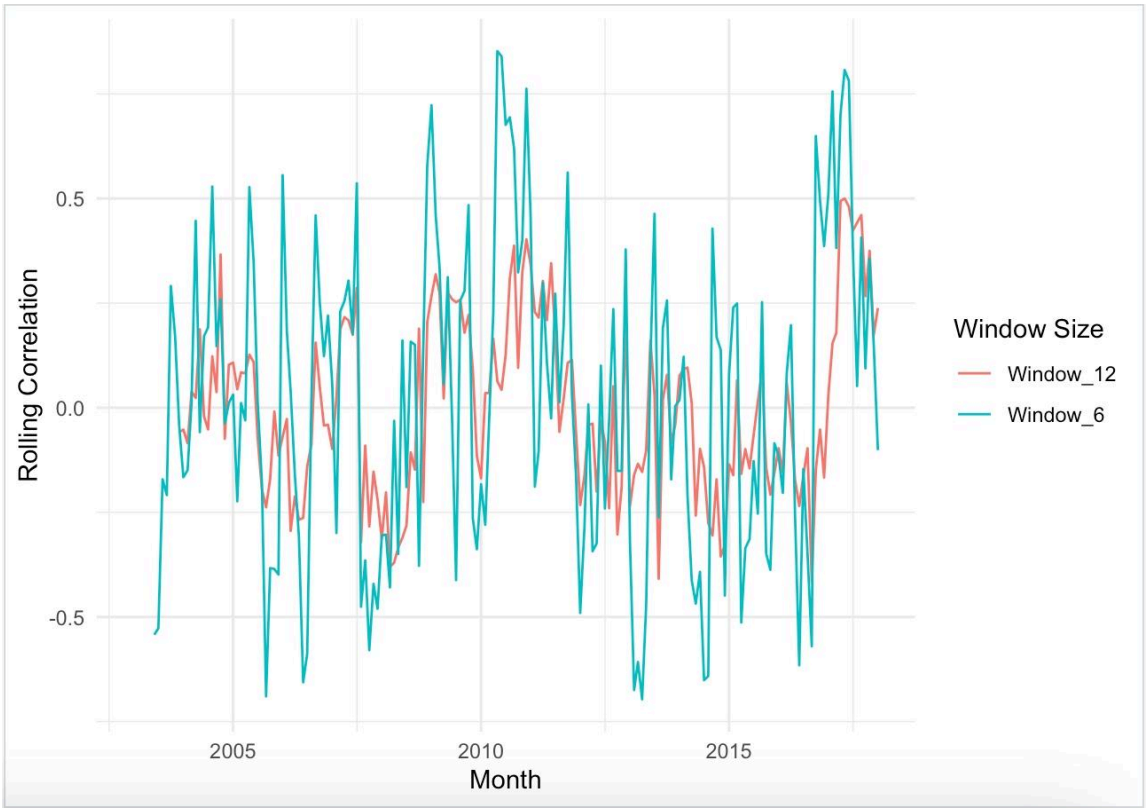

Rolling Correlation (Window Size: 6 months)
1.0
0.5
0.0
-0.5
Correlation
2005
2010
2015
2020
Date





-

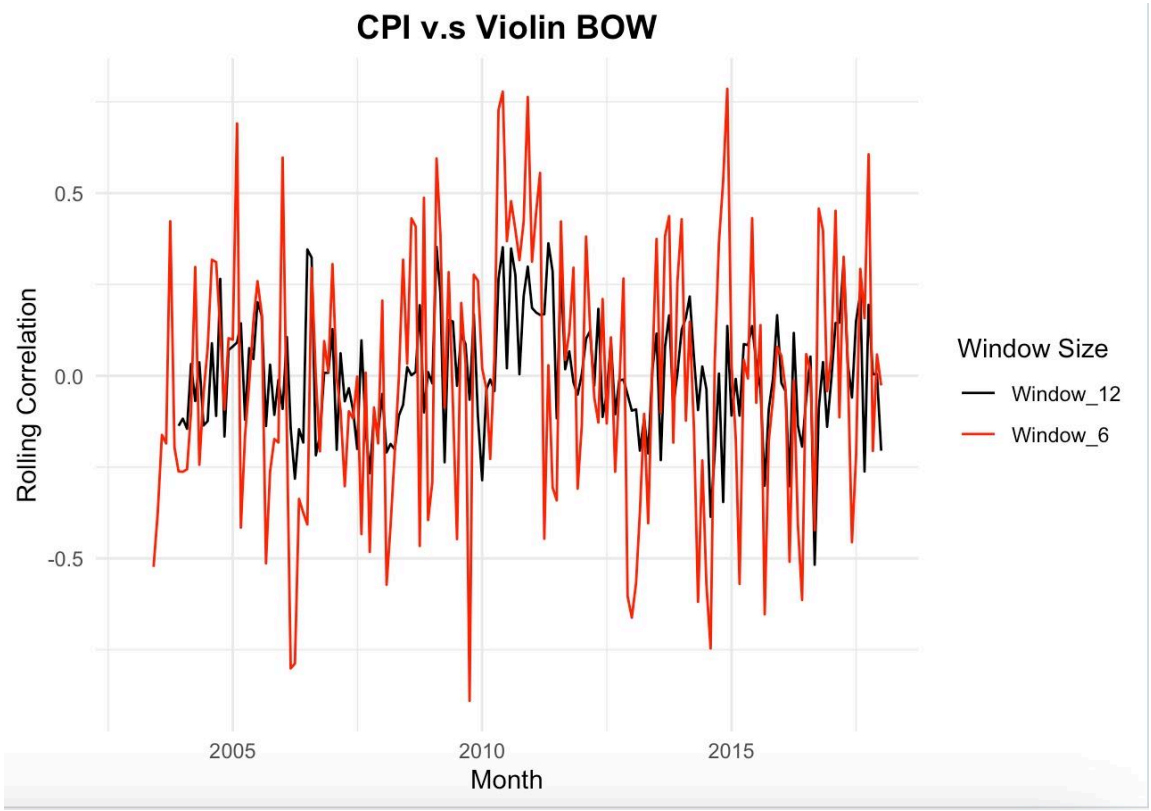
- Label X-axis with specific dates to make it easier to view on the rolling correlation plot-



- Do correlations again with actual CPI inflation rate-

```
> print(correlation_cpi)
                CPIAUCSL monthly_return
CPIAUCSL       1.00000000    -0.02246041
monthly_return -0.02246041     1.00000000
```

-


CPI v.s Violin BOW
Rolling Correlation
0.5
0.0
-0.5
2005
2010
2015
Month
Window Size
Window_12
Window_6