Two directions by paper could go:

1) Probability of a Positive / Negative Earnings Statement based on price changes in the previous (week).

News is hard to predict, random events are random. But a definite report such as a quarterly earnings statement guaranteed to be either positive or negative news, based on whether or not the earnings beats expectations. Statistical prices returns have no known distribution, and are known to have fat tails, compared to a normal distribution.

Weak Hypothesis: Positive stock movement in the time before an earnings report is correlated / an indicator of an earnings that beats (misses) expectations. – which leads to stock price increases (decreases). Another important section of the earnings report is the company’s future outlook. A change in future outlook changes investor models and future expectations.

I want to research on:

* Whether there is increased volatility for stock prices the closer the earnings date.
* Affect on pre-earnings date drift in relation to predicting post earnings
* Whether positive earnings and expected returns are due to increased risk (volatility).
* Using Price & Volume indicator analysis (ex. Bollinger bands, moving averages), show the technical analysis metric on the days leading up to the report
* Whether the earning release date is also a random walk – or whether the change of a positive / negative earnings is 50%
* Conclude with whether a certain finding may be based on insider trading (pessimistic) or simply investor expectations (information -> market efficiency)

Positives: a lot of data, most of research would be using technical analysis & looking at time series

Negatives: determining how much data is enough to find a conclusion, determining the difference between causation, correlation, and risk/return

2) I wanted to pursue a topic related to the Statistics behind Behavioral Finance, Neuroeconomics.

These two topic could potentially come together – I could prospect about how behaviors impact the price changes of a stock before an earnings date.

Positives – perhaps more interesting, more tangible

Negatives – more difficult to obtain data, ex. data of human behavior, but I could use candlestick graphs to measure “herd mentality”

Here are some graphs I have started experimenting with:

\*\*First Quarter\*\*

```{r}

x <- cocacolaQ1

x$Change <- ifelse(x$Close > x$Open, "up", "down")

x$Width <- x$Volume / 10^2

ggplot(x, aes(x=Date))+

 geom\_linerange(aes(ymin=Low, ymax=High)) +

 theme\_bw() +

 labs(title="Coca Cola Candle Chart for First Quarter",

 y = "Stock Value") +

 geom\_rect(aes(xmin = Date - Width/2 \* 0.9,

 xmax = Date + Width/2 \* 0.9,

 ymin = pmin(Open, Close),

 ymax = pmax(Open, Close),

 fill = Change), alpha = 0.6) +

 guides(fill = FALSE, colour = FALSE) +

 scale\_fill\_manual(values = c("down" = "red3", "up" = "limegreen")) +

 geom\_vline(xintercept = as.numeric(cocacolaQ1$Date[32])) +

 annotate("text", label = "First Quarter Earnings",

 x = cocacolaQ1$Date[32], y = 42)

```

