

Homework 3

Ze Yang (zey@andrew.cmu.edu)

Due Thursday, September 21 at 3:00 PM

You should submit the Rmd file with your answers in the appropriate spaces. Rename the file as YOURANDREWID_HW3.Rmd and submit it via Canvas. Also submit the .pdf file that is produced.

Whenever you are asked to create a plot, unless the specific form is stated, you have the flexibility to use your judgement to choose the plot you feel is most appropriate. I expect that you will take steps to make the plot clean and readable.

I am assuming at this point you know how to appropriately specify R commands within this Markdown file, i.e., using the “triple quotation marks.”

1. Go to the following website:

https://www.quandl.com/data/AII/AII_SENTIMENT-AII-Investor-Sentiment-Data and download the data set. (Note that there are multiple ways of doing this, including using the “Download” button in the upper right of the screen.) Read these data into R and create an appropriate data frame. Be sure that each column is in an appropriate form (date, factor, character, etc.) When reading in the file, **do not make any changes to the data file itself**. Resolve any issues using appropriate R commands.

Solution:

```
df = read.csv('AII-AII_SENTIMENT.csv')
df$Date = as.Date(df$Date)
str(df)
```

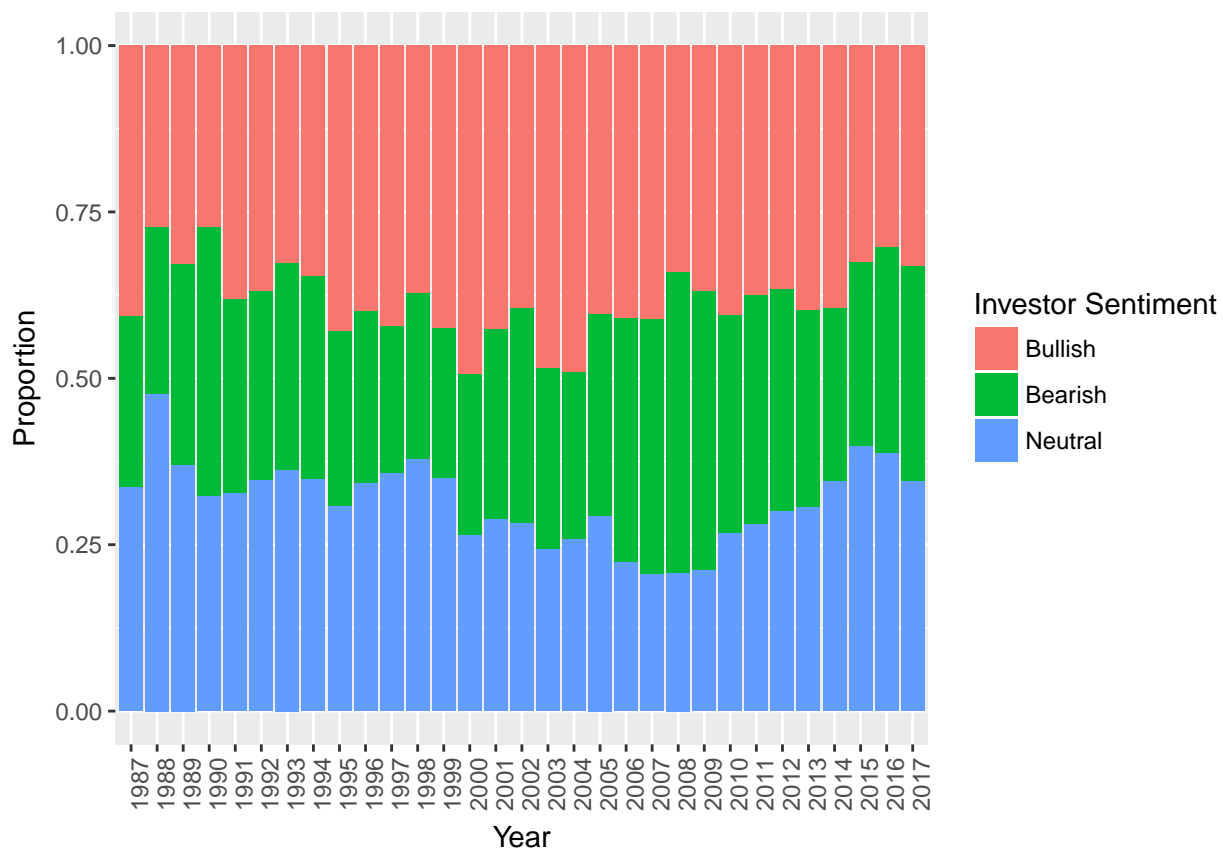
```
## 'data.frame':    1575 obs. of  13 variables:
## $ Date          : Date, format: "2017-09-14" "2017-09-07" ...
## $ Bullish       : num  0.413 0.293 0.25 0.281 0.342 ...
## $ Neutral       : num  0.367 0.35 0.351 0.336 0.33 ...
## $ Bearish       : num  0.22 0.357 0.399 0.383 0.328 ...
## $ Total         : num  1 1 1 1 1 ...
## $ Bullish.8.Week.Mov.Avg : num  0.328 0.32 0.319 0.325 0.327 ...
## $ Bull.Bear.Spread : num  0.1932 -0.0646 -0.1489 -0.1019 0.0138 ...
## $ Bullish.Average : num  0.383 0.383 0.383 0.383 0.383 ...
## $ Bullish.Average...St..Dev : num  0.485 0.485 0.485 0.485 0.485 ...
## $ Bullish.Average...St..Dev.1 : num  0.28 0.28 0.28 0.28 0.28 ...
## $ S.P.500.Weekly.High : num  2498 2480 2460 2475 2475 ...
## $ S.P.500.Weekly.Low  : num  2459 2428 2428 2417 2438 ...
## $ S.P.500.Weekly.Close : num  2498 2466 2458 2444 2468 ...
```

2. Create a figure that shows how the proportion of investors who are bullish/neutral/bearish changes over the **years** of the data set. Do this via a stacked bar chart similar to the one seen on slide 25 in Part 5 of the lecture notes.

Solution:

```
df$Year = format(df$Date, "%Y")
# year aggregation,
# calculate mean sentiment proportion group by year
df.agg.year = df %>% group_by(
  Year=Year) %>% summarize(
  Bullish=mean(Bullish, na.rm=T),
  Bearish=mean(Bearish, na.rm=T),
  Neutral=mean(Neutral, na.rm=T))
df.agg.year = as.data.frame(df.agg.year)
df.agg.year.melt = melt(df.agg.year, id=c('Year'))

# create the plot
ggplot(data=df.agg.year.melt,
  mapping=aes(x=Year, y=value, fill=variable)) +
  geom_bar(position="fill", stat="identity") +
  labs(x="Year", y="Proportion", fill="Investor Sentiment") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



3. A colleague wonders if there could be a seasonal effect, i.e., that there are natural variations in investor sentiment over the course of a year that tend to repeat from year to year. Create a plot that explores whether or not there is evidence for this. Be sure to include a description of your conclusion. **Do not perform any “formal” analyses.**

Solution:

```
month.order = c('Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun',
                'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec')
df$Month = factor(format(df$Date, "%b"), levels = month.order)
# year-month aggregation,
# calculate mean sentiment proportion every year-month
df.agg.ym = df %>% group_by(
  YearMonth=floor_date(Date, "month")) %>% summarize(
  Bullish=mean(Bullish, na.rm=T),
  Bearish=mean(Bearish, na.rm=T),
  Neutral=mean(Neutral, na.rm=T),
  Spread=mean(Bull.Bear.Spread, na.rm=T),
  Close=mean(S.P.500.Weekly.Close, na.rm=T))
# month aggregation,
# calculate mean sentiment proportion group by month
df.agg.month = df %>% group_by(
  Month=Month) %>% summarize(
  Bullish=mean(Bullish, na.rm=T),
  Bearish=mean(Bearish, na.rm=T),
  Neutral=mean(Neutral, na.rm=T),
  Spread=mean(Bull.Bear.Spread, na.rm=T))
df.agg.ym = as.data.frame(df.agg.ym)
df.agg.month = as.data.frame(df.agg.month)
df.agg.month.2 = rbind(df.agg.month, df.agg.month)
df.agg.month.2 = melt(df.agg.month.2, id=c('Month'))
df.agg.month.2$Year = rep(c(1987, 2017), each=12)
df.agg.month.2$Label = "Month Mean"

df.agg.ym.melt = melt(df.agg.ym, id=c('YearMonth'))
df.agg.ym.melt$Year = as.numeric(
  format(df.agg.ym.melt$YearMonth, "%Y"))
df.agg.ym.melt$Month = factor(format(
  df.agg.ym.melt$YearMonth, "%b"), levels = month.order)
df.agg.ym.melt$Label = "YearMonth Mean"
df.agg.ym.melt = df.agg.ym.melt[df.agg.ym.melt$variable!="Close",]

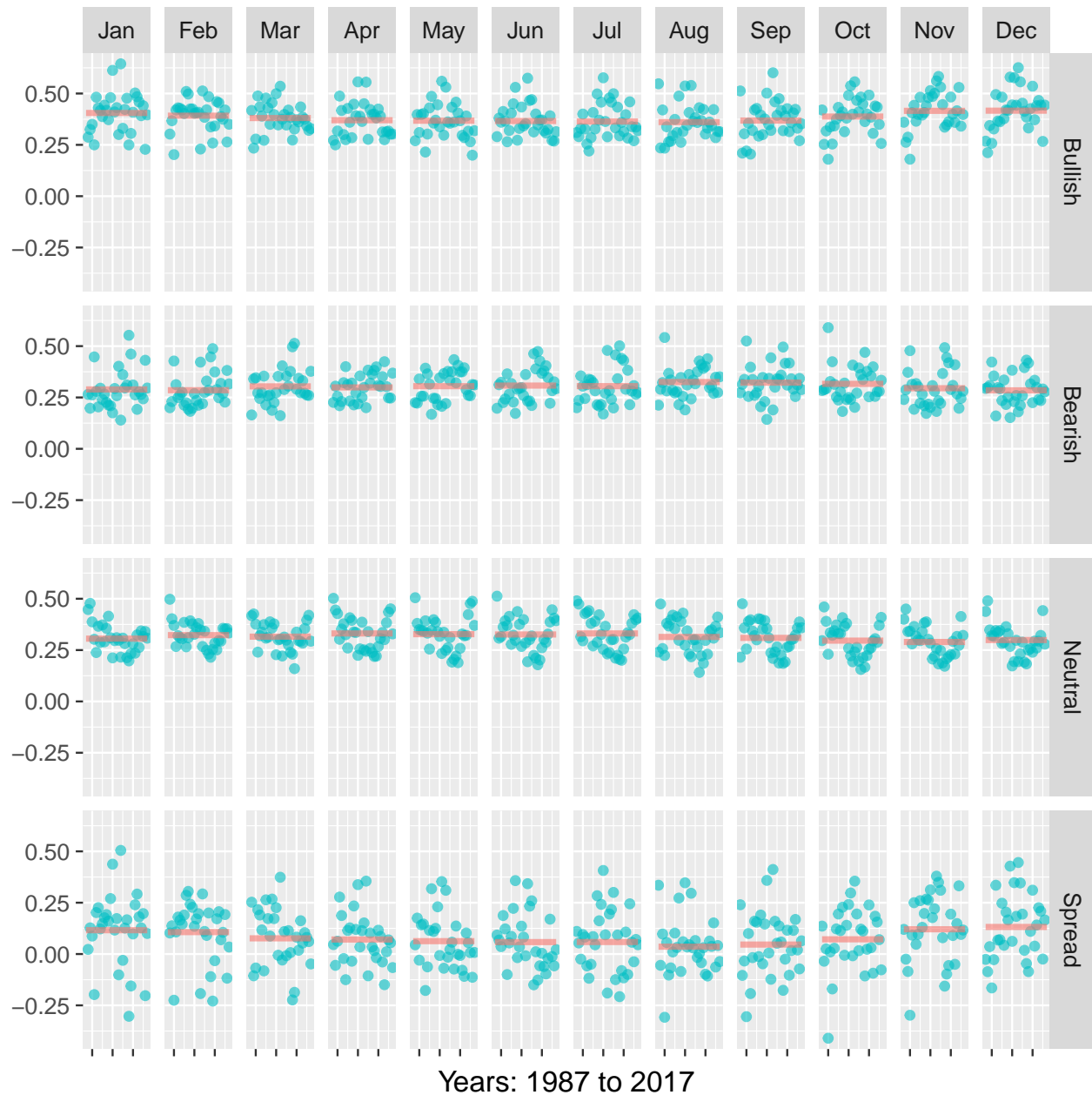
#create the plot
ggplot(df.agg.ym.melt, mapping=aes(x=Year, y=value, colour=Label)) +
  geom_point(data=df.agg.ym.melt,size=I(2),alpha=I(0.6)) +
  geom_line(data=df.agg.month.2,size=I(1.5),alpha=I(0.6)) +
  theme_grey(base_size=15) +
```

```

facet_grid(variable ~ Month) +
theme(legend.title = element_blank(),
      axis.title.y=element_blank(),
      axis.text.x=element_blank(),
      legend.position=c(.1,-.4)) +
xlab("Years: 1987 to 2017") +
ylab("Sentiment Proportion")

```

Warning: Removed 4 rows containing missing values (geom_point).



Conclusion: - The Facets are split by different month groups across X axis, and split by different sentiment variables across Y axis, namely Bullish, Bearish, Neutral and Spread

between bull and bear. - Within each facet subplot, X axis stand for different years. The blue dots represent the **Mean** of sentiment measure of a certain year in a certain month group. The red horizontal line is the mean sentiment measure of that month group aggregated with all year. - If intra-year pattern exists, we expect to see patterns across different monthly groups, i.e. significant difference across different months for each year. However, we didn't observe any intra-year pattern. The monthly aggregates are pretty much the same, and there is a lot of inter-year variations for each group. - We conclude that there is no intra-year seasonal pattern, at least at month level.

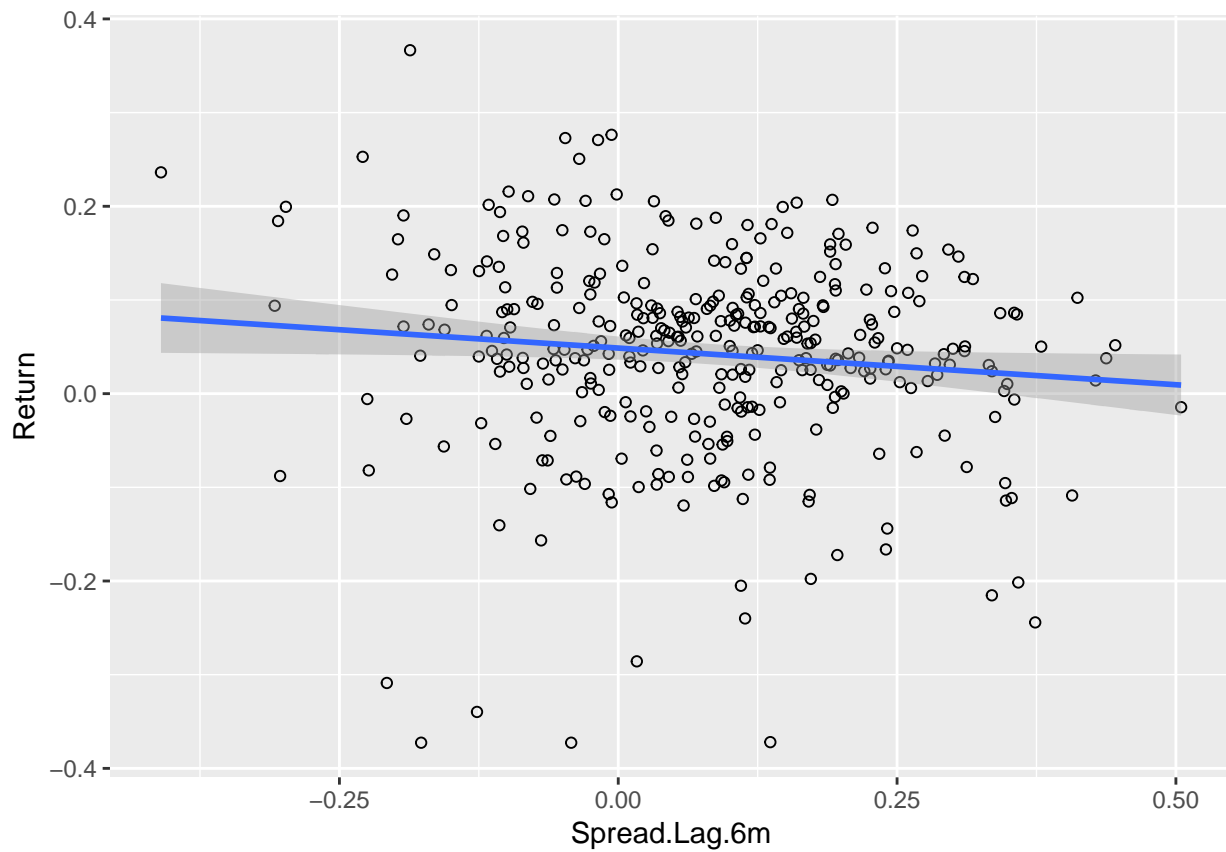
-
4. Do you think that investors are generally “correct” in their sentiments? In other words, does their opinion regarding the future of the market predict where the market will be over the following six months? Create a plot that explores this. Be sure to include a description of your conclusion. **Do not perform any “formal” analyses.**

```
# calculate 6-month return and bull-bear spread before 6 months
df.agg.ym$Spread.Lag.6m = lag(df.agg.ym$Spread, 6)
df.agg.ym$Close.Lag.6m = lag(df.agg.ym$Close, 6)
df.agg.ym$Close.Diff = df.agg.ym$Close - df.agg.ym$Close.Lag.6m
df.agg.ym$Return = df.agg.ym$Close.Diff / df.agg.ym$Close.Lag.6m
```

```
ggplot(df.agg.ym, aes(x=Spread.Lag.6m, y=Return)) +
  geom_point(shape=1) +      # Use hollow circles
  geom_smooth(method=lm)    # Add linear regression line
```

```
## Warning: Removed 7 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 7 rows containing missing values (geom_point).
```



Conclusion: - The X axis stand for Bull-Bear spread. Y axis stand for the future realized S&P index return 6 months from now. - We fail to observe any correlation between X and Y. The scatters are very “randomly” distributed. We conclude that the sentiment cannot predict market movement 6 months from now.