Week 1 Homework

Molly Farry-Thorn

September 4, 2017

1. Move your data into a long format and a wide format. Did you have any specific challenges that you encountered? If so, discuss them.

```
letters <- read.delim("Masters coded letters.csv", header=TRUE, sep=",")
long_to_wide <- unite(letters, "Participant", c("Subject","Speaker"))

long_to_wide <- gather(long_to_wide, Var, Val, select=c("Age", "Length", "Utterances", "Utterances.with
    unite(VarG, Var, Session) %>%
    spread(VarG, Val)
long_to_wide <- long_to_wide[,c(1:14,18:25,15:17,26:169)]
```

2. Create a wave variable and date variable (if applicable).

I have a variable for the session number and a variable for the child's exact age. I don't have the dates of the sessions

3. What is your sample size for each wave of assessment?

```
ddply(letters, .(Session,Speaker), summarize, N = length(Speaker))
```

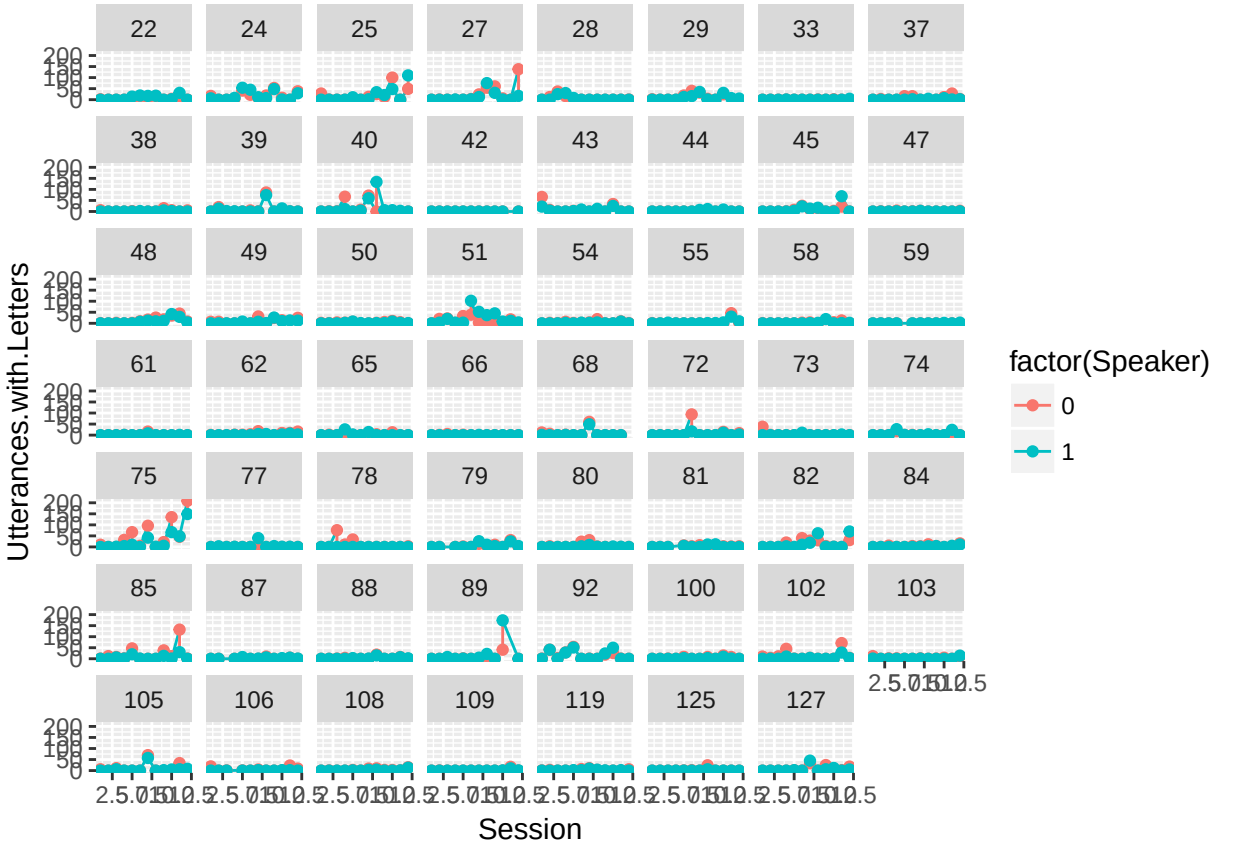
```
##    Session Speaker  N
## 1        1       0 55
## 2        1       1 55
## 3        2       0 55
## 4        2       1 55
## 5        3       0 53
## 6        3       1 53
## 7        4       0 53
## 8        4       1 53
## 9        5       0 54
## 10       5       1 54
## 11       6       0 55
## 12       6       1 55
## 13       7       0 55
## 14       7       1 55
## 15       8       0 55
## 16       8       1 55
## 17       9       0 54
## 18       9       1 55
## 19      10       0 54
## 20      10       1 55
## 21      11       0 51
## 22      11       1 53
## 23      12       0 54
## 24      12       1 54
```

4. Take the date variable and convert it to a different date format such as time in study or age (if appropriate). What scale is most suitable for your analyses? (weeks/months/years?)

I don't have the dates of the sessions

5. Graph your data using the different time metrics, fitting individual curves for each person.

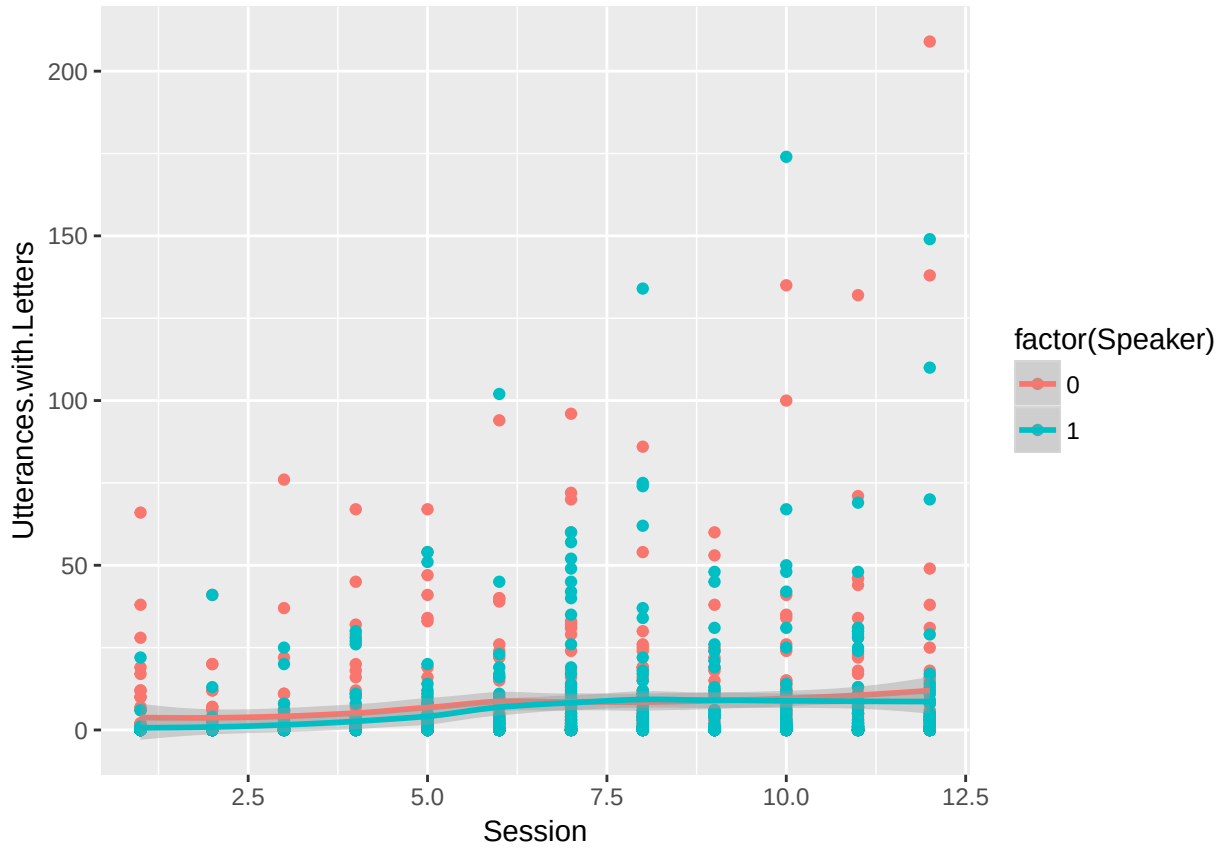
```
gg1 <- ggplot(letters,
   aes(Session, Utterances.with.Letters, group = Subject, color = factor(Speaker))) + geom_line() +  ge
gg1
```



6. Create an overall average trend of your data (split up into groups if appropriate). Attempt to color your individual data points and/or shade different lines (highlight some particiapnts, highlight the average trend line but not the individual level lines)

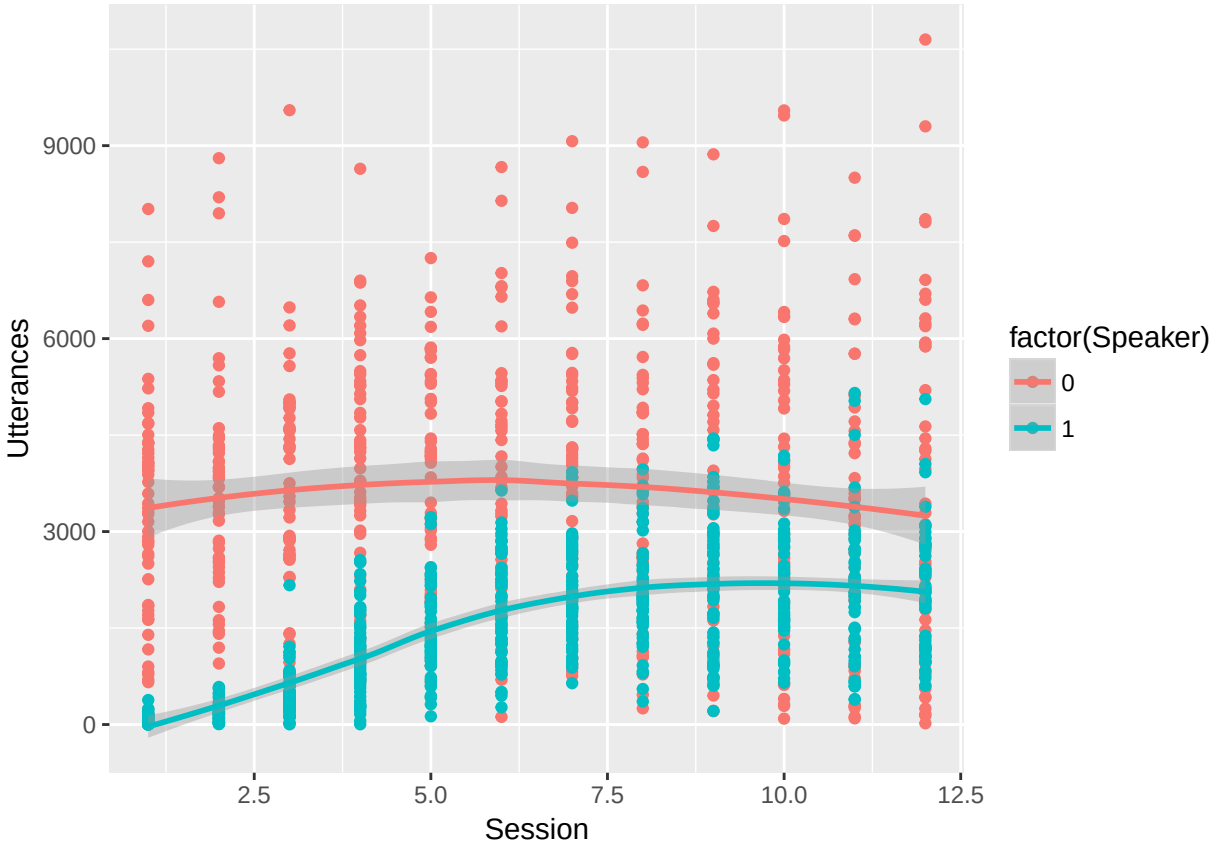
```
gg2 <- ggplot(letters, aes(Session, Utterances.with.Letters, color = factor(Speaker))) +
       geom_point() + geom_smooth()
gg2
```

```
## `geom_smooth()` using method = 'loess'
```




```
gg3 <- ggplot(letters, aes(Session, Utterances, color = factor(Speaker))) +
       geom_point() + geom_smooth()
gg3
```

```
## `geom_smooth()` using method = 'loess'
```



7. Look at the correlations of your DV across time

```
func <- function(letters)
{return(data.frame(CORR = cor(letters$Utterances, letters$Utterances.with.Letters)))}
ddply(letters,.(Session), func)
```

```
##    Session      CORR
## 1        1 0.1978969
## 2        2 0.1868895
## 3        3 0.1389265
## 4        4 0.2990457
## 5        5 0.3335335
## 6        6 0.1126958
## 7        7 0.3056780
## 8        8 0.0724001
## 9        9 0.4338104
## 10      10 0.2581181
## 11      11 0.4253248
## 12      12 0.3898306
```