Landa-LDA assignment 1

Isidro Landa

September 4, 2017

Load packages

```r
library(tidyverse)
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr

## Conflicts with tidy packages ----------------------------------------------

## filter(): dplyr, stats
## lag():    dplyr, stats
```

```r
library(haven)    # For some reason not loading with tidyverse
library(magrittr)
```

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

## The following object is masked from 'package:purrr':
##
##     set_names

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

```r
library(ggthemes)
library(plyr)
```

```
## -------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)

## -------------------------------------------------------------------------

##
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

## The following object is masked from 'package:purrr':
##
##     compact
```

```
rm(list=ls())
```

Load data

```
## Set working directory
setwd("C:/Users/Isidro/OneDrive/WUSTL/Coursework/F17/ALDA")

long <- read_sav("Landa-CC daily diary.sav")
```

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.

```
## Long to wide
long$day <- as.numeric(long$day)
wide <- long %>%
  spread(day, curraffect)
wide
```

```
## # A tibble: 1,050 x 93
##       id   URM  URM2  day0 homeavoid.x e.excited e.enthus e.nerv e.fear
##    * <dbl> <dbl> <dbl> <dbl>       <dbl>     <dbl>    <dbl>  <dbl>  <dbl>
## 1     1     0     0     0          NA        NA       NA     NA     NA
## 2     1     0     0     1          NA        NA       NA     NA     NA
## 3     1     0     0     2        1.50         3        3      1      1
## 4     1     0     0     3        1.50         4        4      1      1
## 5     1     0     0     4          NA        NA       NA     NA     NA
## 6     1     0     0     5        1.25         2        2      3      2
## 7     1     0     0     6        1.25         2        2      3      2
## 8     2     0     1     0          NA        NA       NA     NA     NA
## 9     2     0     1     1        1.25         1        1      2      2
## 10    2     0     1     2        1.25         3        2      2      1
## # ... with 1,040 more rows, and 84 more variables: e.hostile <dbl>,
## #   e.uncertain <dbl>, e.calm <dbl>, e.peace <dbl>, e.indiff <dbl>,
## #   e.homesick <dbl>, e.guilty <dbl>, e.happy <dbl>, e.shame <dbl>,
## #   e.slug <dbl>, e.sleepy <dbl>, e.proud <dbl>, neg.m.x <dbl>,
## #   pos.m.x <dbl>, er.accept <dbl>, er.distractbusy <dbl>, er.rumin <dbl>,
## #   er.prob <dbl>, er.socsupp <dbl>, er.benfind <dbl>, er.conseq <dbl>,
## #   er.selfblame <dbl>, er.otherblame <dbl>, er.posfocus <dbl>,
## #   er.beh <dbl>, er.savor <dbl>, er.perspect <dbl>,
## #   er.distractonline <dbl>, er.suppress1 <dbl>, er.suppress2 <dbl>,
## #   er.avoid <dbl>, iof.day <dbl>, WUcon_hrs <dbl>, WUcon_mode <dbl>,
## #   WUcon_mode_txt <dbl>, WUcon_mode_3 <dbl>, WUcon_mode_4 <dbl>,
## #   WUcon_mode_5 <dbl>, WUcon_mode_5_txt <chr>, famcon <dbl+lbl>,
## #   fambff_modes_1 <dbl>, fambff_modes_2 <dbl>, fambff_modes_3 <dbl>,
## #   fambff_modes_4 <dbl>, fambff_modes_5 <dbl>, fambff_modes_5_txt <chr>,
## #   famcon_me <dbl>, me_famcon <dbl>, interact <dbl>, express <dbl+lbl>,
## #   suppress <dbl+lbl>, avoid.home <dbl>, avoid.h_acad <dbl>,
```

```
## #   avoid.h_soc <dbl>, avoid.h_money <dbl>, avoid.college <dbl>,
## #   avoid.c_acad <dbl>, avoid.c_soc <dbl>, avoid.c_money <dbl>,
## #   firstgen <dbl+lbl>, gender <dbl+lbl>, iof.day.y <dbl>, neg.m.y <dbl>,
## #   pos.m.y <dbl>, homeavoid.y <dbl>, er.avoid.y <dbl>,
## #   `filter_$` <dbl+lbl>, min <dbl+lbl>, missincases <dbl+lbl>,
## #   completecases <dbl>, complete <dbl+lbl>, iof.x.pc <dbl>,
## #   duration.exp <dbl>, duration <dbl>, StartDate <chr>, EndDate <chr>,
## #   Date_analysis <chr>, `1` <dbl>, `2` <dbl>, `3` <dbl>, `4` <dbl>,
## #   `5` <dbl>, `6` <dbl>, `7` <dbl>
```

```
#
# test <- long %>%
#   select(id, day, curraffect, e.excited)
#
# widetest <- test %>%
#   spread(day, e.excited)
```

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

```
## Already included.
```

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

```
ddply(long, .(day), summarize, N = length(day))
```

```
##   day   N
## 1   1 150
## 2   2 150
## 3   3 150
## 4   4 150
## 5   5 150
## 6   6 150
## 7   7 150
```

```
ddply(long, .(day, er.avoid), summarize, N = length(er.avoid))
```

```
##    day er.avoid  N
## 1    1        1 45
## 2    1        2 34
## 3    1        3 24
## 4    1        4 10
## 5    1        5  3
## 6    1       NA 34
## 7    2        1 64
## 8    2        2 35
## 9    2        3 23
## 10   2        4 10
## 11   2        5  1
## 12   2       NA 17
## 13   3        1 54
## 14   3        2 47
```

```
## 15   3        3 20
## 16   3        4  8
## 17   3       NA 21
## 18   4        1 36
## 19   4        2 30
## 20   4        3 14
## 21   4        4  2
## 22   4        5  3
## 23   4       NA 65
## 24   5        1 49
## 25   5        2 38
## 26   5        3 21
## 27   5        4  2
## 28   5        5  3
## 29   5       NA 37
## 30   6        1 55
## 31   6        2 35
## 32   6        3 17
## 33   6        4  4
## 34   6        5  3
## 35   6       NA 36
## 36   7        1 53
## 37   7        2 41
## 38   7        3 12
## 39   7        4  6
## 40   7        5  2
## 41   7       NA 36
```

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?)

```
# Created a var (duration.exp) to exclude 'extreme' values of duration in completing the
# survey from the time participant started. Already included.

# 1 = duration was roughly within what was expected (i.e., 2-10 min)
# Intended duration was 5-7 min
```

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

```
g1 <- ggplot(long, aes(day, er.avoid, group = id)) + geom_line() +
  geom_point() + facet_wrap(~id) + theme_classic() +
  xlab("Day") + ylab("Emotion Regulation-Avoidance") +
  ggtitle("Avoidance as a Funcation of Day")
g1
```

```
## Warning: Removed 115 rows containing missing values (geom_path).
```

```
## Warning: Removed 246 rows containing missing values (geom_point).
```
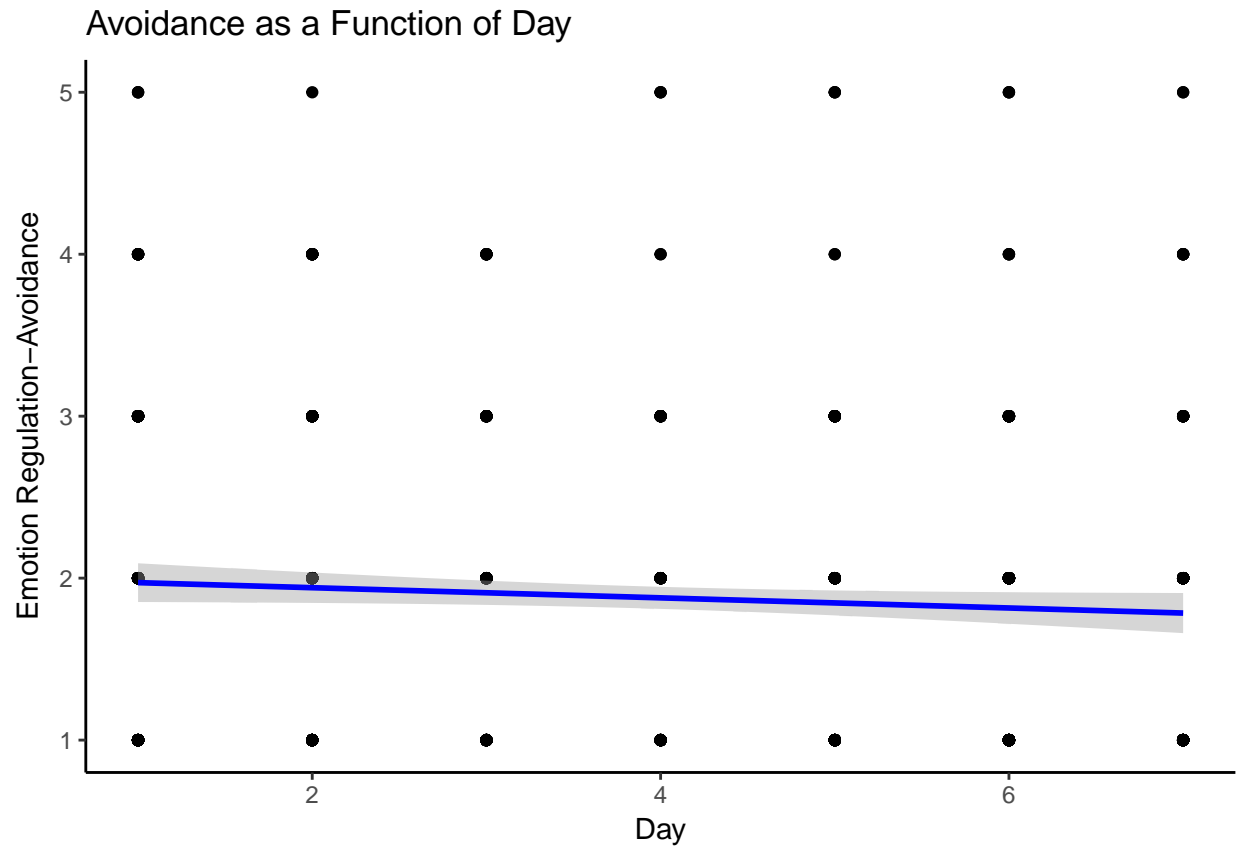
Avoidance as a Funcation of Day
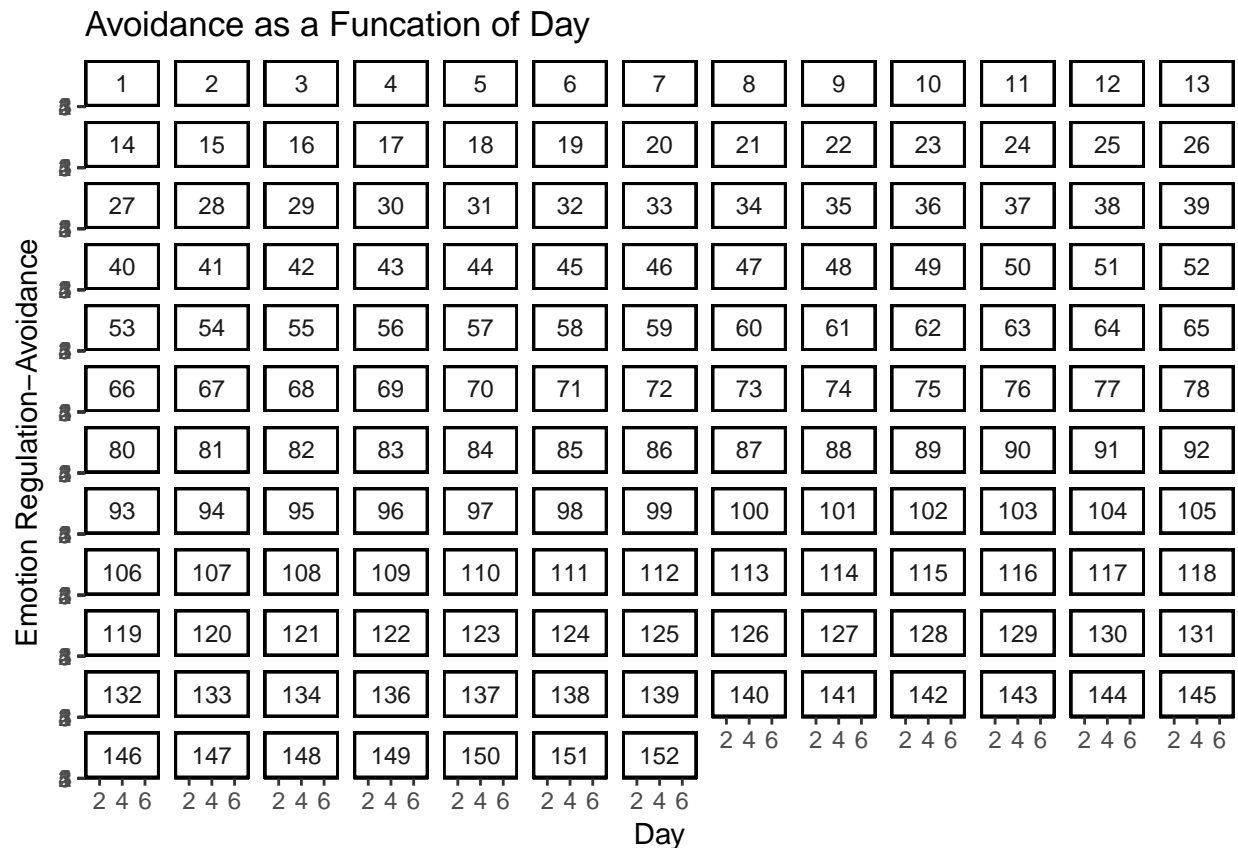
6. Create an overall average trend of your data (split up into groups if appropriate). Attempt to color your individual data points and/or shade different lines (highlight some particiapnts, highlight the average trend line but not the individual level lines)

```r
# Overall trend in avoidance
g2 <- ggplot(long, aes(day, er.avoid)) + geom_point() +
  stat_smooth(method = "lm", color = "blue") + theme_classic() +
  xlab("Day") + ylab("Emotion Regulation-Avoidance") +
  ggtitle("Avoidance as a Function of Day")
g2
```

```
## Warning: Removed 246 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 246 rows containing missing values (geom_point).
```

Avoidance as a Function of Day

```
# Line for and colored by each participant
g3 <- ggplot(long, aes(day, er.avoid, group = id)) +
  geom_point() + stat_smooth(method = "lm", se = F) +
  aes(color = factor(id)) + guides(color = F) +
  theme_classic() +
  xlab("Day") + ylab("Emotion Regulation-Avoidance") +
  ggtitle("Avoidance as a Function of Day")
g3
```

Warning: Removed 246 rows containing non-finite values (stat_smooth).

Warning: Removed 246 rows containing missing values (geom_point).

Avoidance as a Function of Day
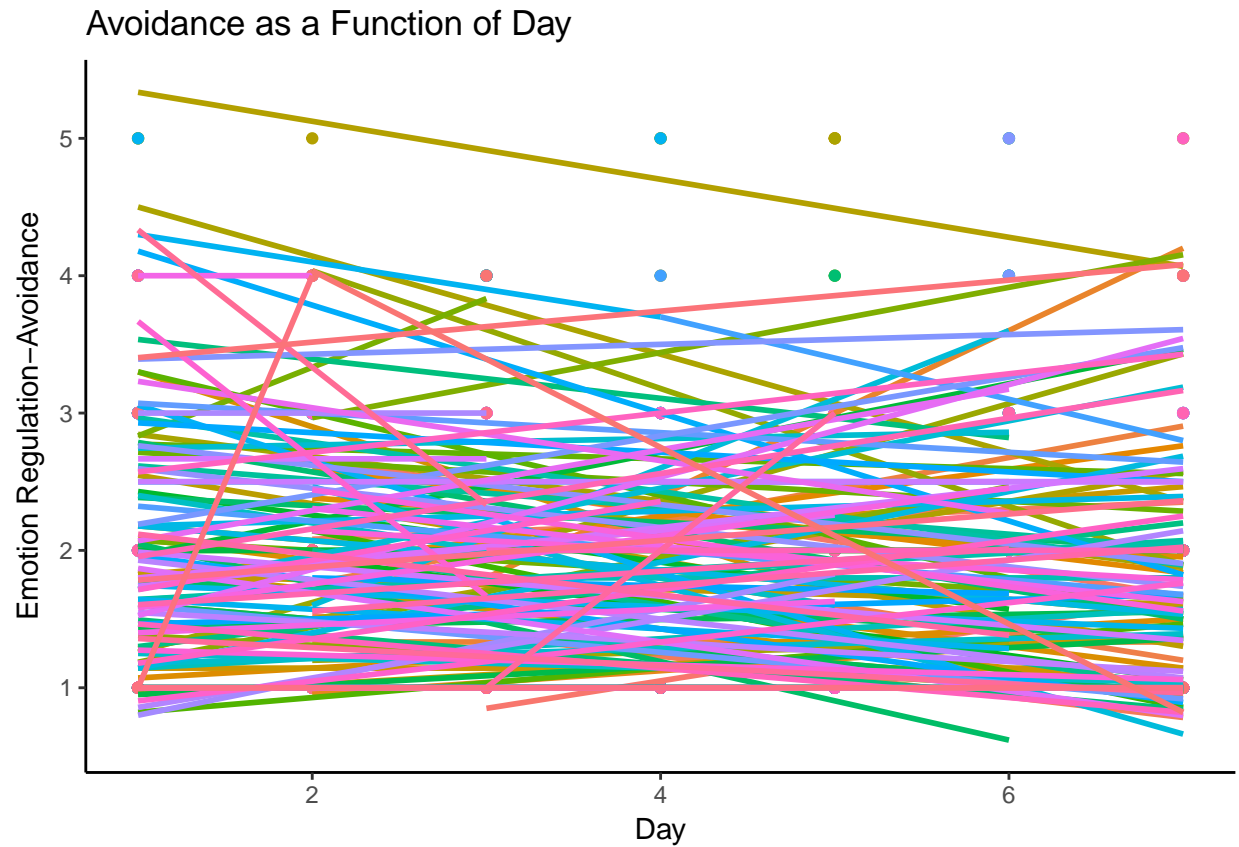

```
# g4 <- ggplot(long, aes(day, er.avoid, group = id)) +
#   stat_smooth(method = "lm", se = F) +
#   theme_classic() +
#   xlab("Day") + ylab("Emotion Regulation-Avoidance") +
#   ggtitle("Avoidance as a Function of Day") +
#   stat_summary(aes(group = 1), fun.y = mean) +
#   stat_summary(aes(group = 1), method = "lm",
#                color = blues9)
# g4
```

7. Look at the correlations of your DV across time