Basic Data Visualization with ggplot2

John D. Lee and Linda Ng Boyle

10/28/2019

ggplot2 for Data Visualization

– skimr provides a simple summary of data

– ggplot2 as a grammar of graphics:

- Basic scatterplot
- Faceting for small multiples
- Loess and linear regression layers
- Reordering for easy comparisons
  – *Typical repeated measures data: "sleepstudy"*
- Reaction (reaction time)
- Days (day of participation)
- Subject (participant ID)

– Does sleep deprivation increase reaction times for people? in a uniform way?

Useful Packages

- `library(skimr)` for skim
- `library(tidyverse)` for ggplot2, dplyr

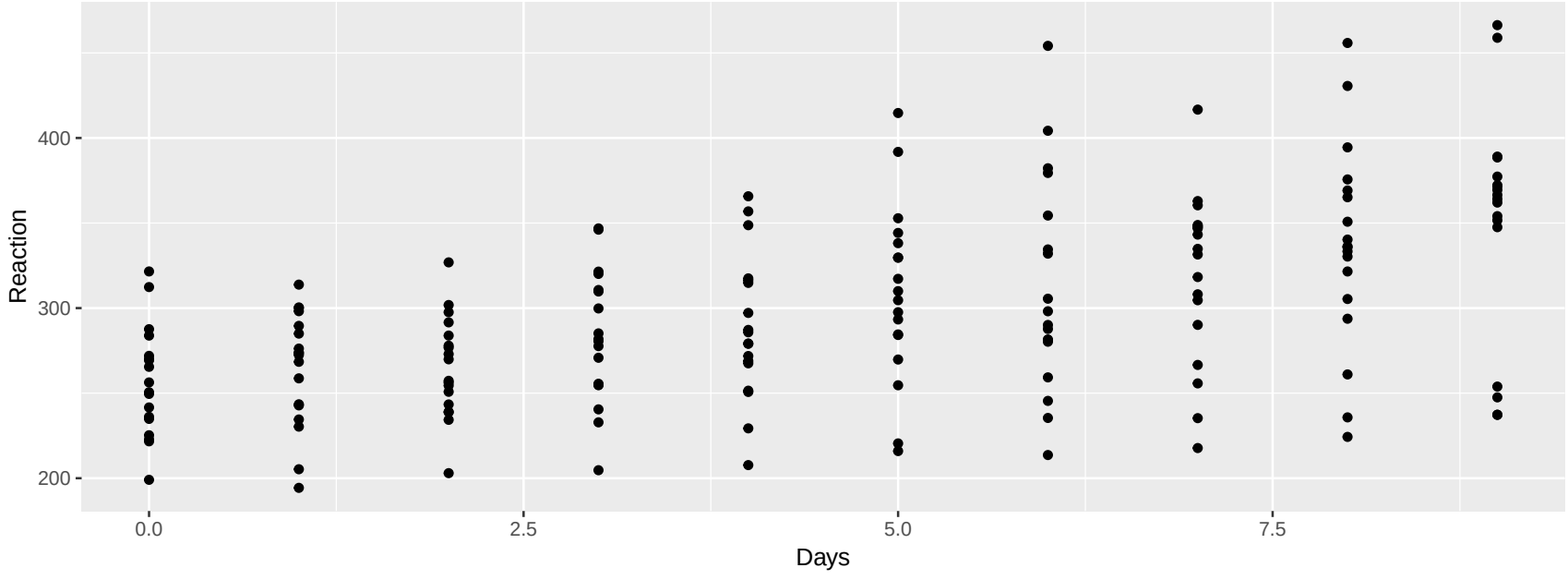
skimr to Summarize Data

```
library(skimr)
library(lme4) # For the sleep study data
sleep.df = sleepstudy
#skim(sleep.df)
```

Simple Scatterplot|Mapping variables to aesthetic properties of geometric elements

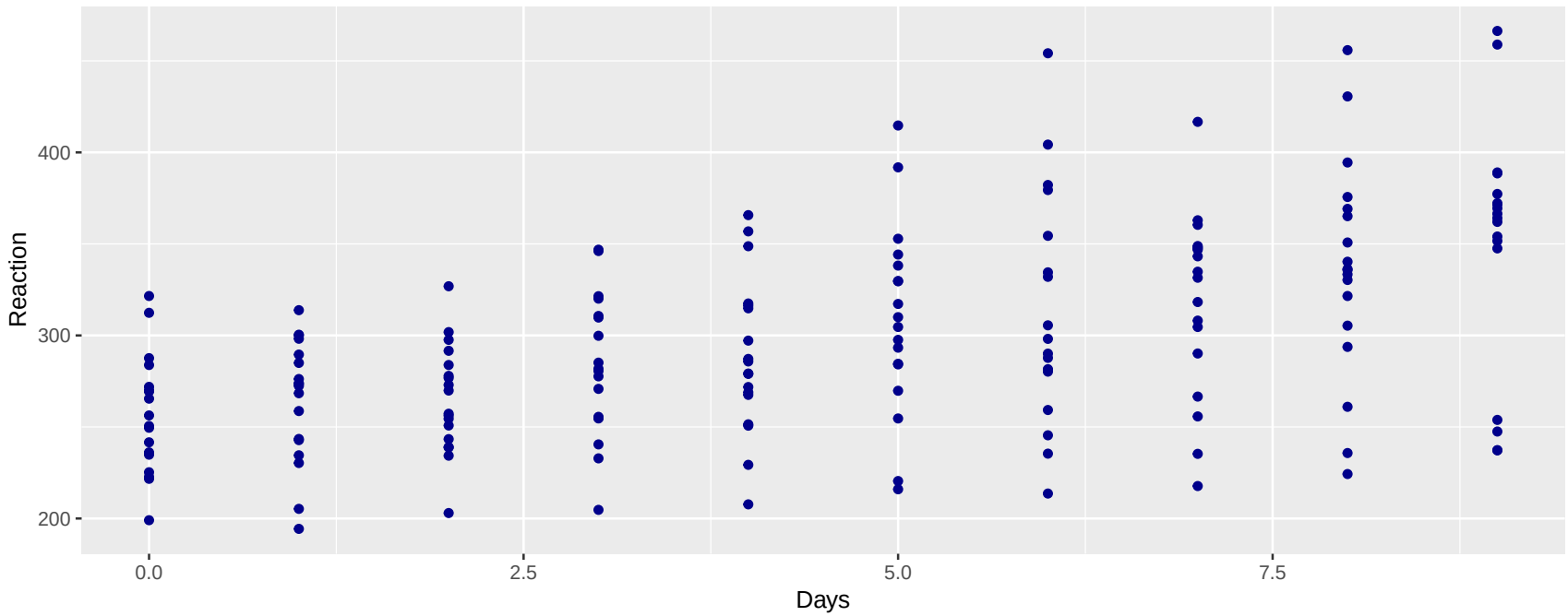
```
library(ggplot2)

ggplot(data = sleep.df, mapping = aes(x = Days, y = Reactio
  geom_point()
```



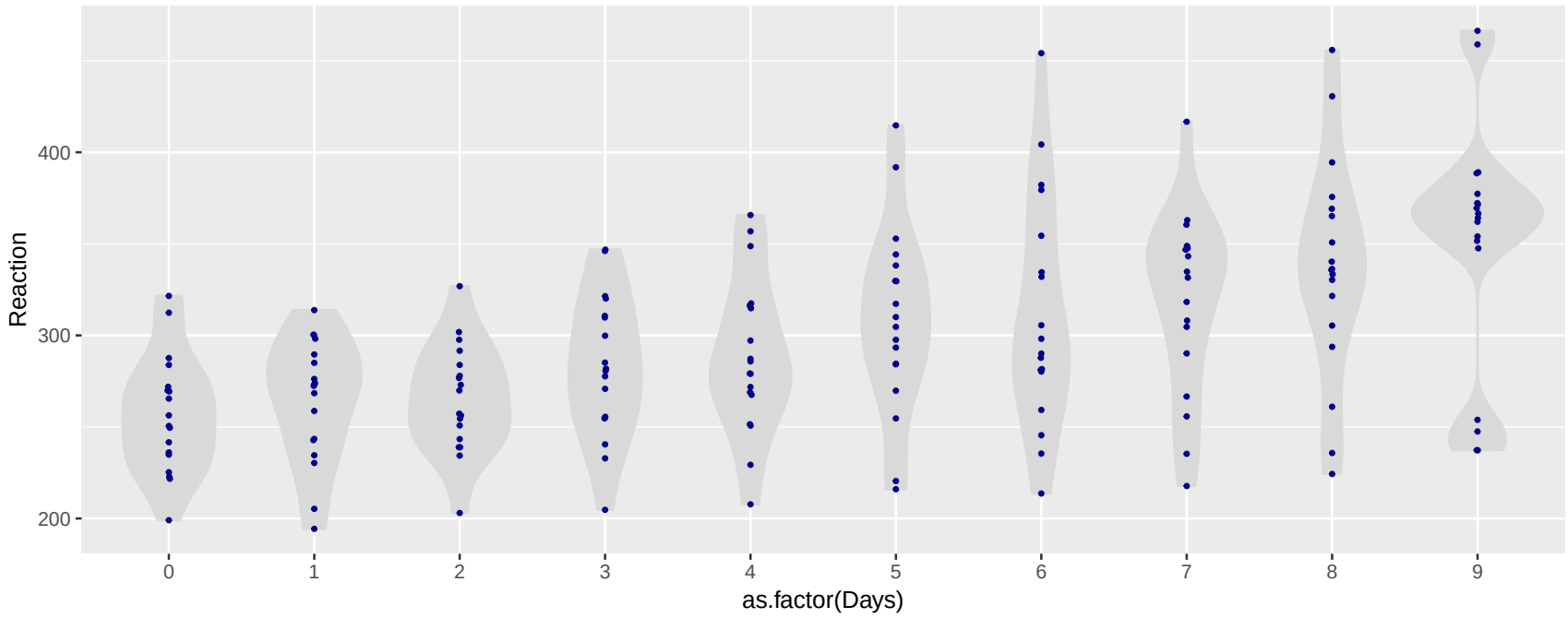
Simple Scatterplot|Setting aesthetic properties

```
ggplot(data = sleep.df, mapping = aes(x = Days, y = Reactio
  geom_point(colour = "darkblue")
```



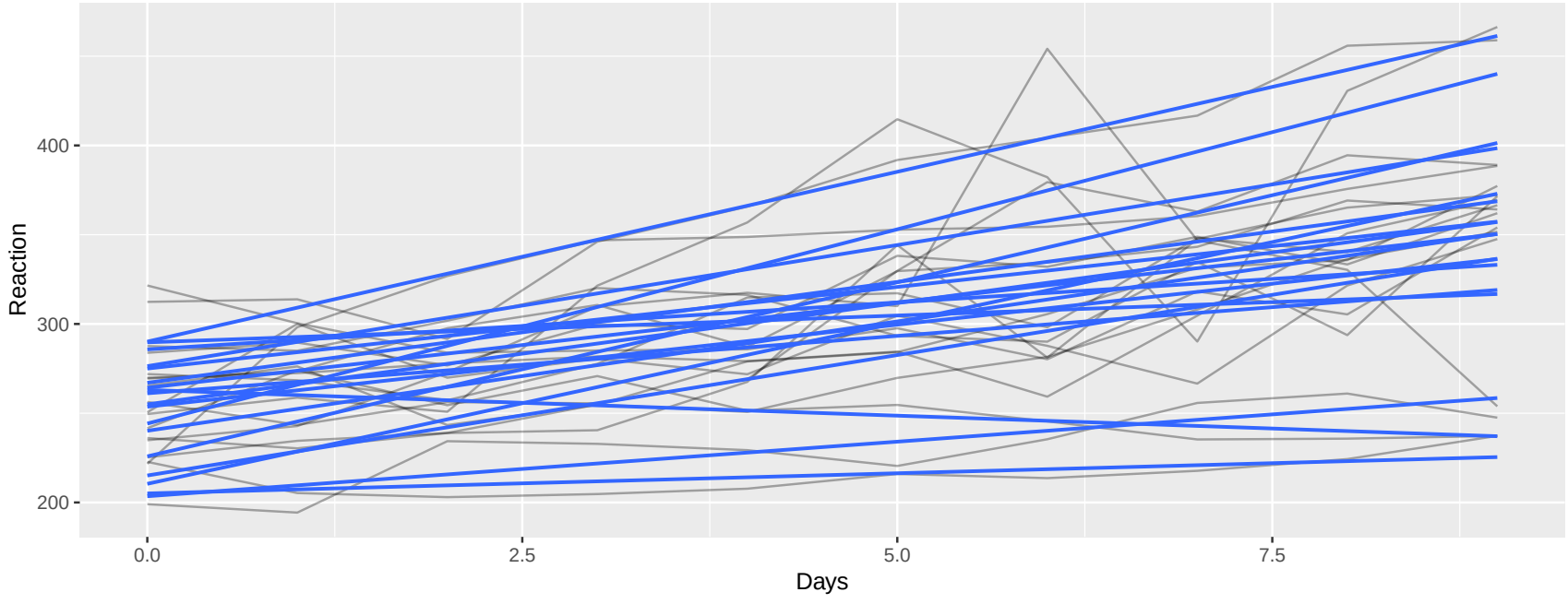
Simple Scatterplot|Layering geometric elements

```
ggplot(data = sleep.df, mapping = aes(x = as.factor(Days),
  geom_violin(fill= "grey85", colour = "grey85") +
  geom_sina(colour = "darkblue", size = .7)
```



Grouped Scatterplot with Linear Regression|Slopes and intercepts both vary across people

```
ggplot(sleep.df, aes(Days, Reaction, group = Subject)) +
  geom_line(alpha = .33) +
  geom_smooth(method = "lm", se = FALSE, size = .8)
```




```
## What happens when you don't identify the groups?
# ggplot(sleep.df, aes(Days, Reaction)) +
#   geom_point() +
#   geom_smooth(method = "lm", se = FALSE)
```

Exercise|Simple Scatterplot

–Replicate what we just showed in an R notebook – Create a Notebook: File>>New file>> R Notebook

– Load the lmer, skimr, and tidyverse packages

– Create the dataframe `sleep.df` by setting it equal to `sleepstudy`

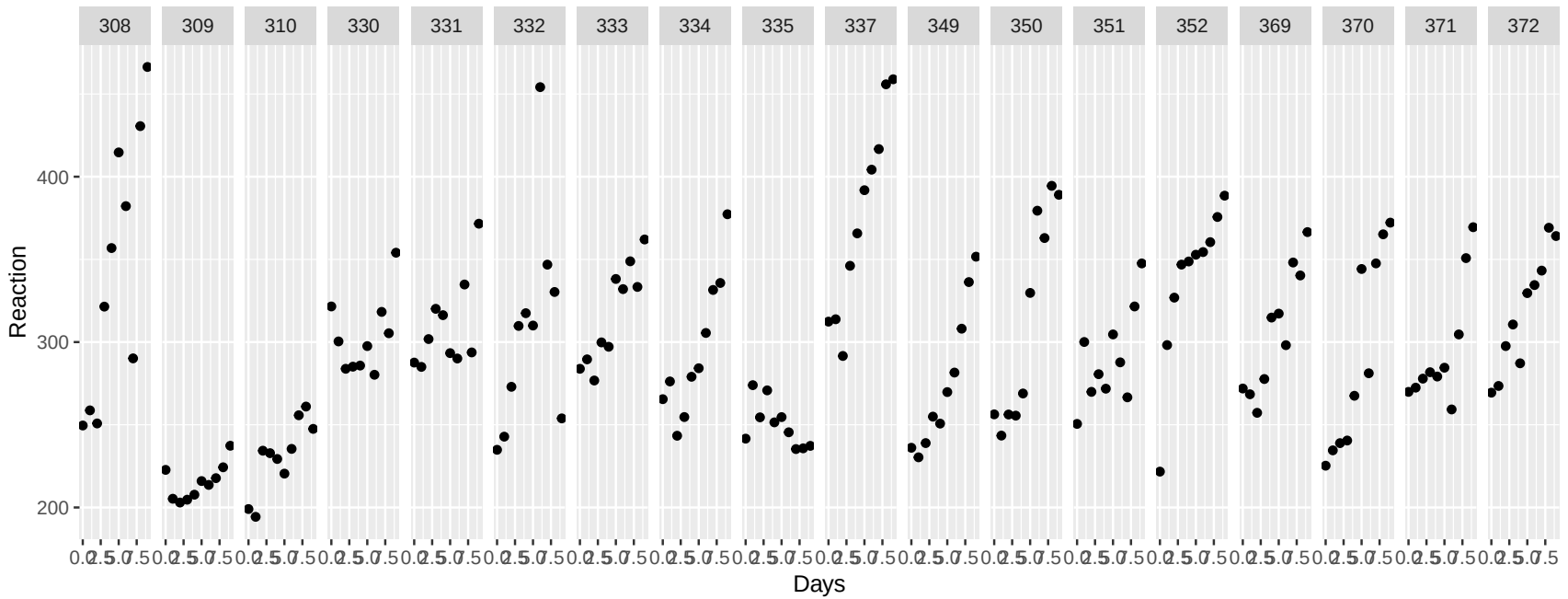
– Use `skimr` to summarize the data

– Use ggplot2 to create simple scatterplot of reaction time and day, and set the color to `darkred`

– Divide code into blocks and add text to make understandable

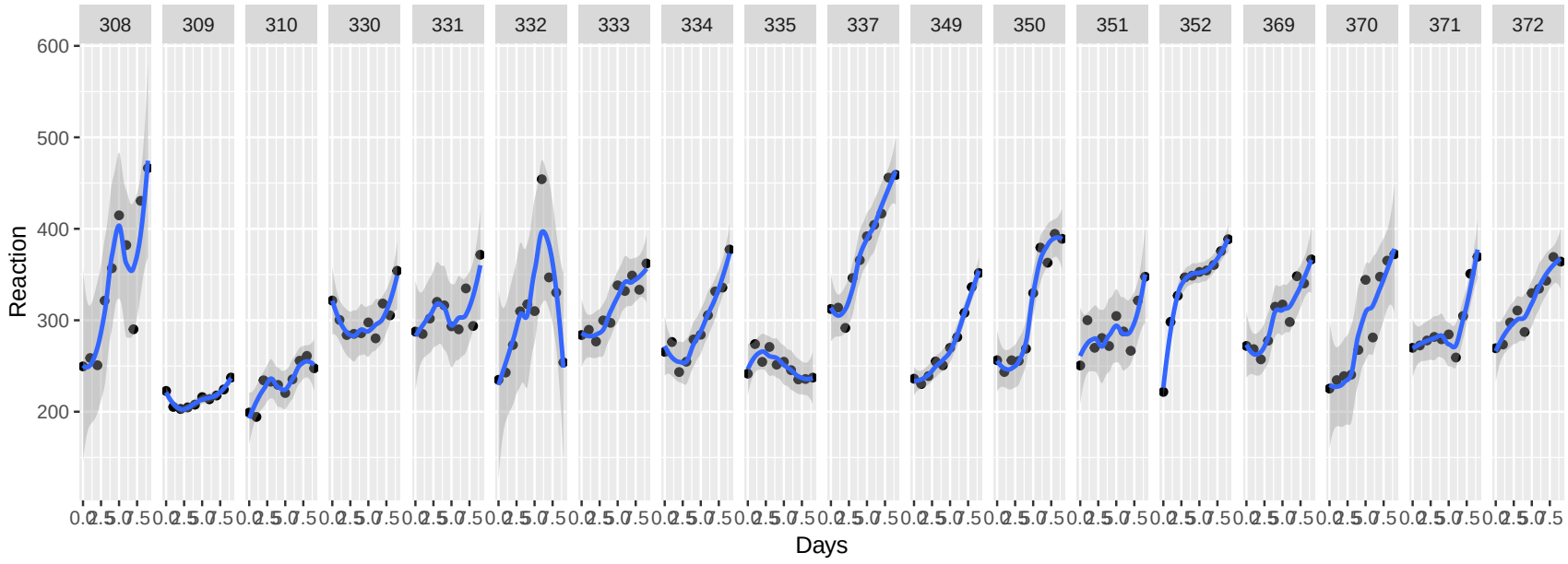
Faceted Scatterplot|Small multiples to show individual responses

```
ggplot(data = sleep.df, mapping = aes(x = Days, y = Reactio
  geom_point() +
  facet_grid(.~Subject)
```



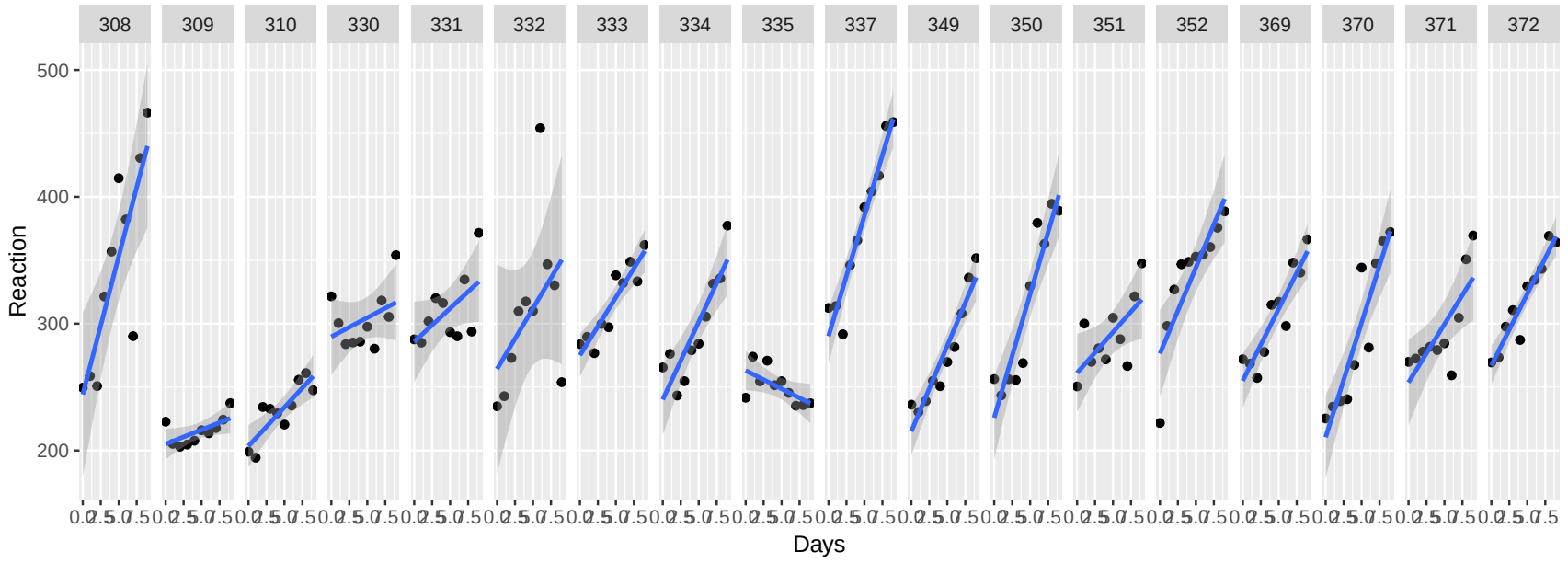
Faceted Scatterplot with Loess Fit|A layer of loess fits shows a linear model doesn't fit all participants

```
ggplot(sleep.df, aes(Days, Reaction)) +
  geom_point() +
  geom_smooth(method = "loess") +
  facet_grid(.~Subject)
```



Faceted Scatterplot with Linear Regression|A linear model shows slopes and intercepts differ

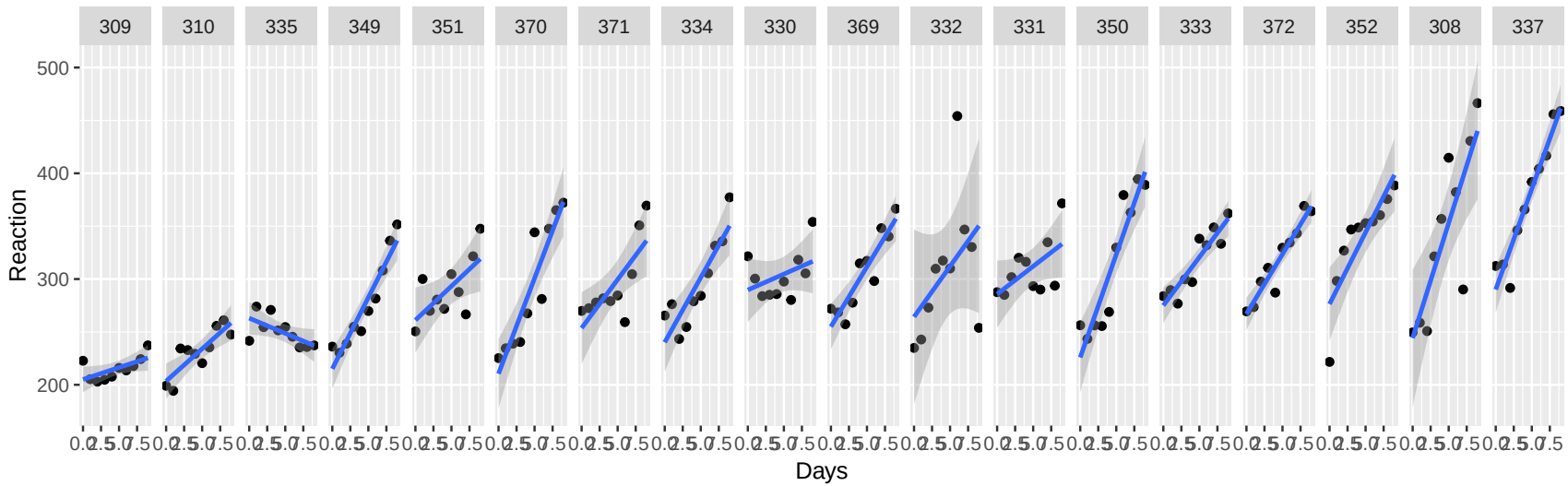
```
ggplot(sleep.df, aes(Days, Reaction)) +
  geom_point() +
  geom_smooth(method = "lm") +
  facet_grid(.~Subject)
```



Faceted Scatterplot with Linear Regression|Subjects ordered by mean reaction time

```
sleep.df = sleep.df %>%  mutate(Subject = reorder(Subject,

ggplot(sleep.df, aes(Days, Reaction)) +
  geom_point() +
  geom_smooth(method = "lm") +
  facet_grid(.~Subject)
```



Exercise|Replicate the faceted scatterplot by adding it to your R notebook

- Create a scatterplot faceted by `Subject`
- Try `+ facet_grid(.~Subject)` and `+ facet_grid(Subject~.)`
- Add a layer of `+ geom_smooth()`
- Order Subjects by standard deviation of `Reaction`

Visualization: A critical first step for any modeling

- Loess and linear regression layers
- Faceting for small multiples
- Reordering for easy comparisons