```
setwd("~/Dropbox/Classes/Longitudinal Data Analysis")
wgt <- read.table("weightslong.csv", sep = ",", header = TRUE)
library(tidyr)

## Warning: package 'tidyr' was built under R version 3.3.2

library(dplyr)

## Warning: package 'dplyr' was built under R version 3.3.2
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

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

library(lubridate)

##
## Attaching package: 'lubridate'
## The following object is masked from 'package:plyr':
##
##     here
## The following object is masked from 'package:base':
##
##     date

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.3.2

library(lme4)

## Loading required package: Matrix
##
## Attaching package: 'Matrix'
## The following object is masked from 'package:tidyr':
##
##     expand
```

1 Question 1

Run linear models on all of your subjects (a basic regression). What is the average intercept, the average slope?

```
colnames(wgt)[3] <- "week"
regression <- lm(weight ~ week, data = wgt)
summary(regression)
```

```
##
## Call:
## lm(formula = weight ~ week, data = wgt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -93.733 -34.377  -0.967  24.023 206.373
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 234.72342    1.39792  167.91   <2e-16 ***
## week         -0.79835    0.06585  -12.12   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 48.51 on 3898 degrees of freedom
##   (1799 observations deleted due to missingness)
## Multiple R-squared:  0.03633,Adjusted R-squared:  0.03609
## F-statistic:   147 on 1 and 3898 DF,  p-value: < 2.2e-16
```

The average intercept is 234.7, which is the average weight of all individuals across time. The average slope is -.798, which is the average weight loss per week across participants.

2 Question 2

Now run a mlm/lmer model with only a random intercept. What is the ICC? What does residual variance look like compared to linear model? Create a graph to show this effect.

```
model1 <- lmer(weight ~ 1 + (1 | ID), data = wgt)
summary(model1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: weight ~ 1 + (1 | ID)
##    Data: wgt
##
## REML criterion at convergence: 29522.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9151 -0.5917 -0.1142  0.5112  4.8113
```

```
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 2465.02  49.649
##  Residual              89.98   9.486
## Number of obs: 3900, groups:  ID, 139
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  224.445      4.215   53.25
```

```
ICC <- 2456/(2465 + 89.98)
ICC
```

```
## [1] 0.96126
```

```
# much more variance between than within people The residual variance in the
# hierarchical model is much smaller
library(sjPlot)
```

```
## Warning:  package 'sjPlot' was built under R version 3.3.2
## Error:  package or namespace load failed for 'sjPlot'
```

```
sjp.lmer(model1, facet.grid = FALSE, sort = "sort.all")
```

```
## Error in eval(expr, envir, enclos):  could not find function "sjp.lmer"
```

```
# NOTE: can't get SJP package to load right
```

3 Question 3

Introduce a fixed slope term. What is the difference in terms of the fixed effects estimates between this estimate and the previous? Of the residual standard error? Create a graph to show both fixed effects estimates and the CIs around them.

```
model2 <- lmer(weight ~ 1 + week + (1 | ID), data = wgt)
summary(model2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: weight ~ 1 + week + (1 | ID)
##    Data: wgt
##
## REML criterion at convergence: 26156.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.1539 -0.5597 -0.1263  0.4739  5.8045
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
```

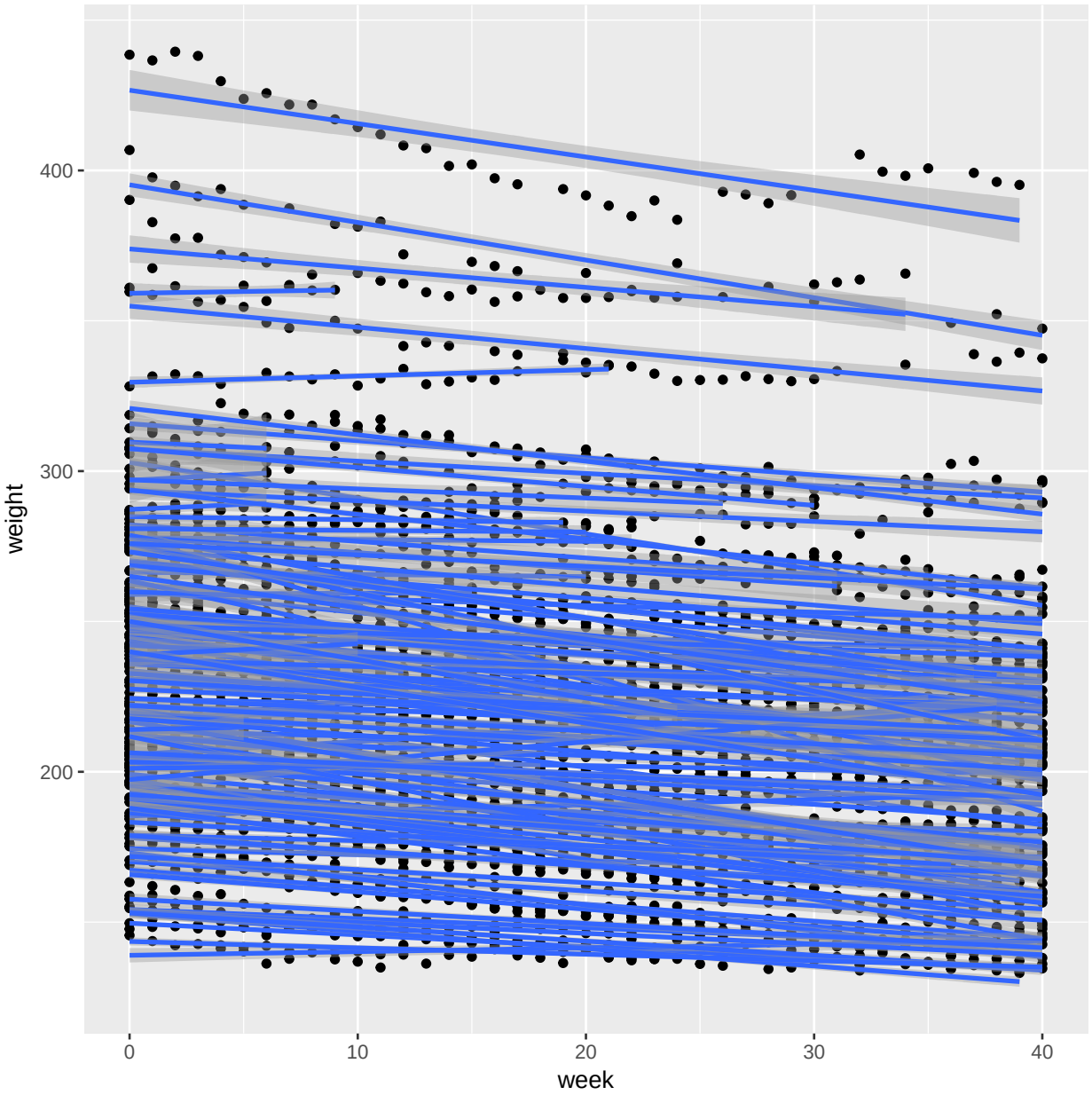
```
##  ID       (Intercept) 2395.87  48.948
##  Residual              36.75   6.062
## Number of obs: 3900, groups:  ID, 139
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 234.720596   4.155461   56.48
## week         -0.633317   0.008576  -73.84
##
## Correlation of Fixed Effects:
##      (Intr)
## week -0.033
```

```
summary(model1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: weight ~ 1 + (1 | ID)
##    Data: wgt
##
## REML criterion at convergence: 29522.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9151 -0.5917 -0.1142  0.5112  4.8113
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 2465.02  49.649
##  Residual               89.98   9.486
## Number of obs: 3900, groups:  ID, 139
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  224.445      4.215   53.25
```

```
gg <- ggplot(wgt, aes(x = week, y = weight, group = ID)) + geom_point() + stat_smooth(method = "lm")
gg
```

```
## Warning:  Removed 1799 rows containing non-finite values (stat_smooth).
## Warning:  Removed 1799 rows containing missing values (geom_point).
```



In this model, the intercept increased. There is also now a "week" fixed effect. The residual of the random effects decreased substantially.

4 Question 4

Run an additional model with a random slope. How does this change compare to the previous model? Should you keep the random slope or not?

```
model3 <- lmer(weight ~ 1 + week + (1 + week | ID), data = wgt)
summary(model3)

## Linear mixed model fit by REML ['lmerMod']
```

```
## Formula: weight ~ 1 + week + (1 + week | ID)
##    Data: wgt
##
## REML criterion at convergence: 23130.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.6267 -0.5689 -0.0626  0.4829  5.2640
##
## Random effects:
##  Groups   Name        Variance  Std.Dev. Corr
##  ID       (Intercept) 2540.0619 50.3990
##           week           0.1698  0.4121  -0.22
##  Residual               14.4718  3.8042
## Number of obs: 3900, groups:  ID, 139
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 234.40100    4.27643   54.81
## week         -0.58689    0.03638  -16.13
##
## Correlation of Fixed Effects:
##      (Intr)
## week -0.219
```

```
summary(model2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: weight ~ 1 + week + (1 | ID)
##    Data: wgt
##
## REML criterion at convergence: 26156.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.1539 -0.5597 -0.1263  0.4739  5.8045
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 2395.87  48.948
##  Residual               36.75   6.062
## Number of obs: 3900, groups:  ID, 139
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept) 234.720596   4.155461   56.48
## week         -0.633317   0.008576  -73.84
##
## Correlation of Fixed Effects:
##      (Intr)
## week -0.033
```

```
anova(model2, model3)

## refitting model(s) with ML (instead of REML)

## Data: wgt
## Models:
## model2: weight ~ 1 + week + (1 | ID)
## model3: weight ~ 1 + week + (1 + week | ID)
##        Df   AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
## model2  4 26162 26187 -13077    26154
## model3  6 23143 23180 -11565    23131 3023.1      2  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Adding week as a random effect reduces the residual of the random effect, although it also reduces the effect of the fixed effect. This is because some of the weekly variance is now being explained within person. It is beneficial to the model to keep the random slope in.

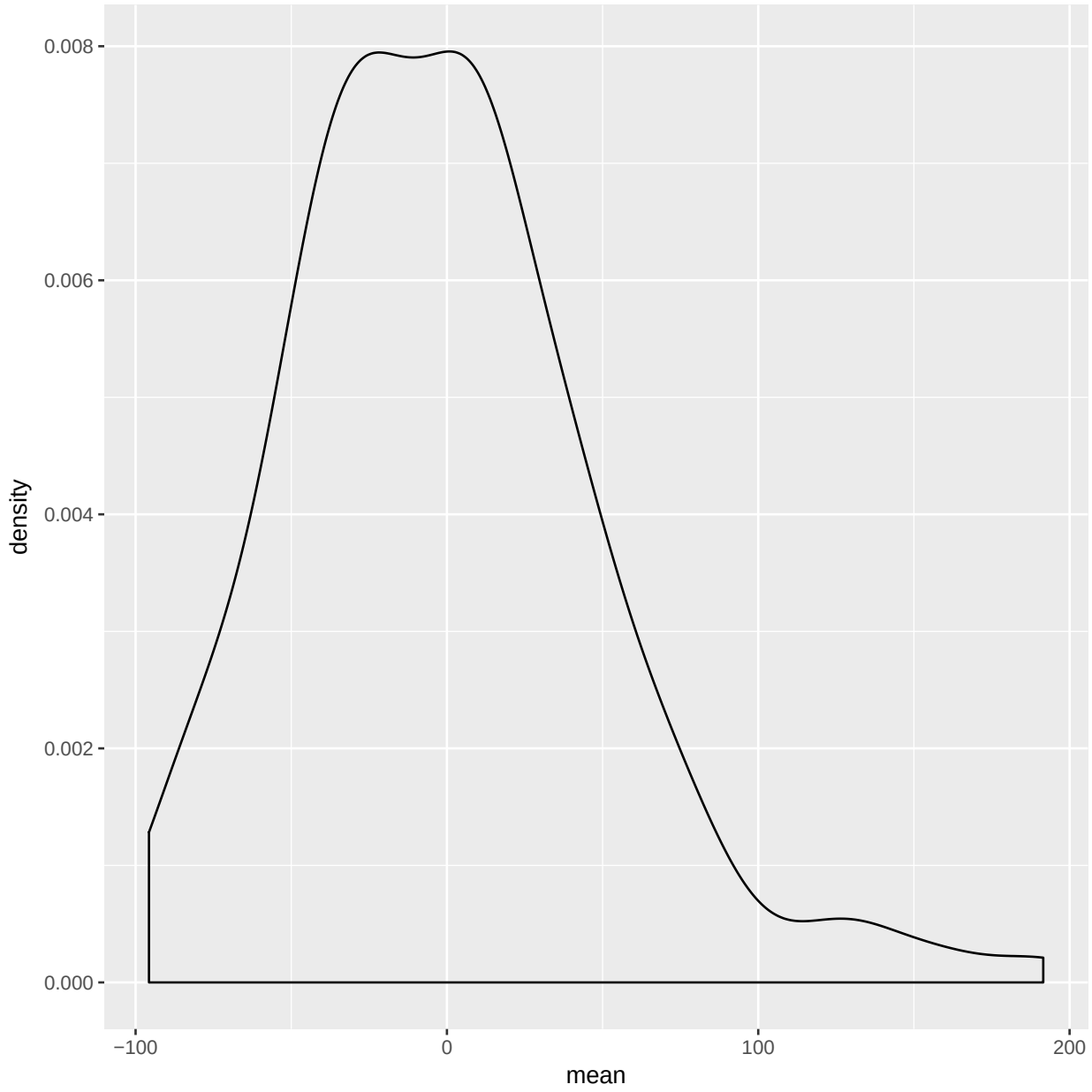
5 Question 5

Interpret the correlation between the slope and the intercept.
If an individual's intercept is higher at the beginning of the program than they will lose weight faster

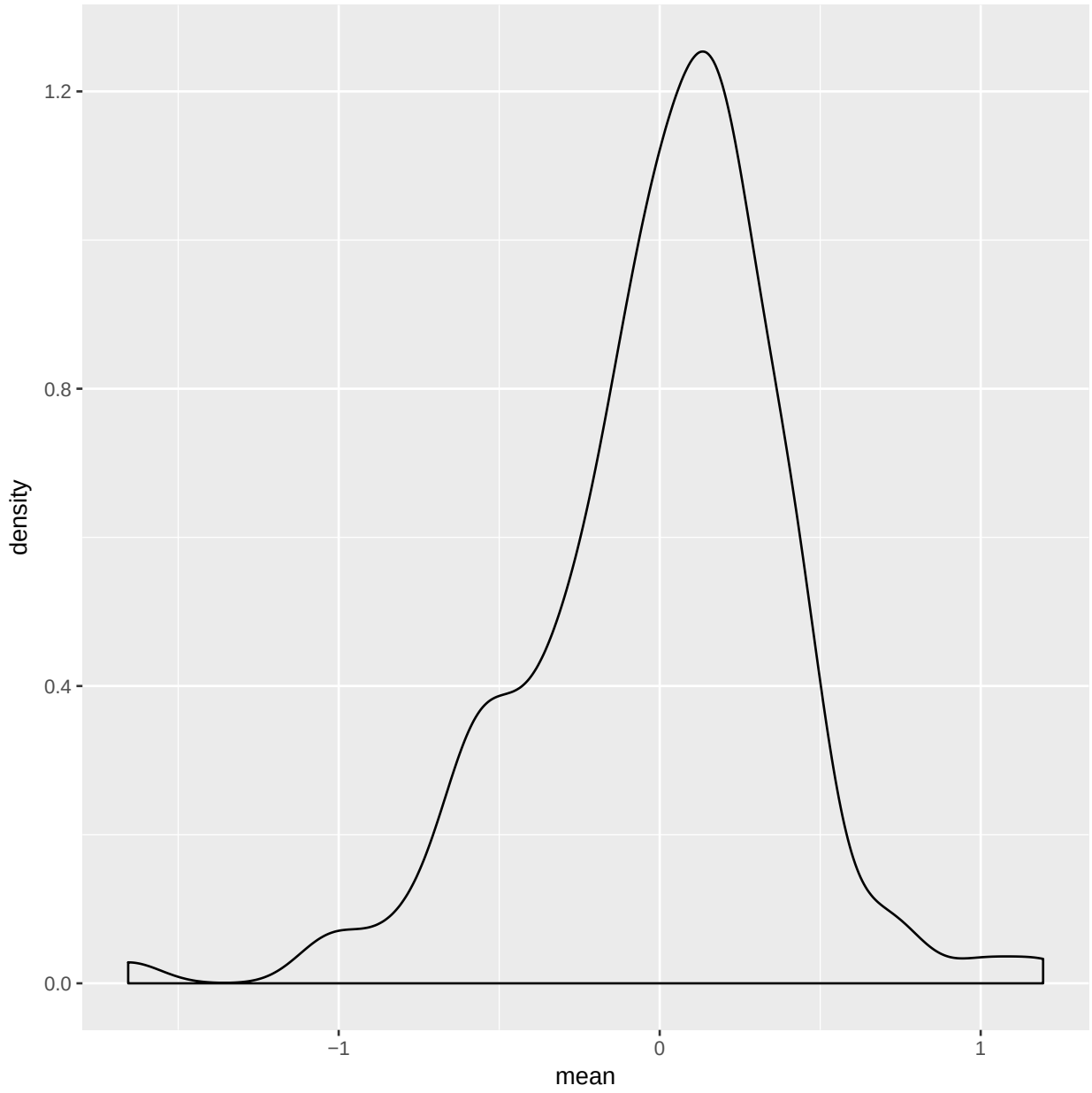
6 Question 6

Create a density plot of the random effects from your final model.

```
library(merTools)

## Warning: package 'merTools' was built under R version 3.3.2
## Loading required package: arm
## Loading required package: MASS
##
## Attaching package: 'MASS'
## The following object is masked from 'package:dplyr':
##
##     select
##
## arm (Version 1.9-1, built: 2016-8-21)
## Working directory is /Users/jacquelinehayes/Dropbox/Classes/Longitudinal Data Analysis

head(re.sim)

## Error in head(re.sim): object 're.sim' not found

re.sim <- REsim(model3)

## Warning: package 'bindrcpp' was built under R version 3.3.2

p1.gg <- re.sim %>% filter(term == "(Intercept)")
ggplot(p1.gg, aes(mean)) + geom_density()
```



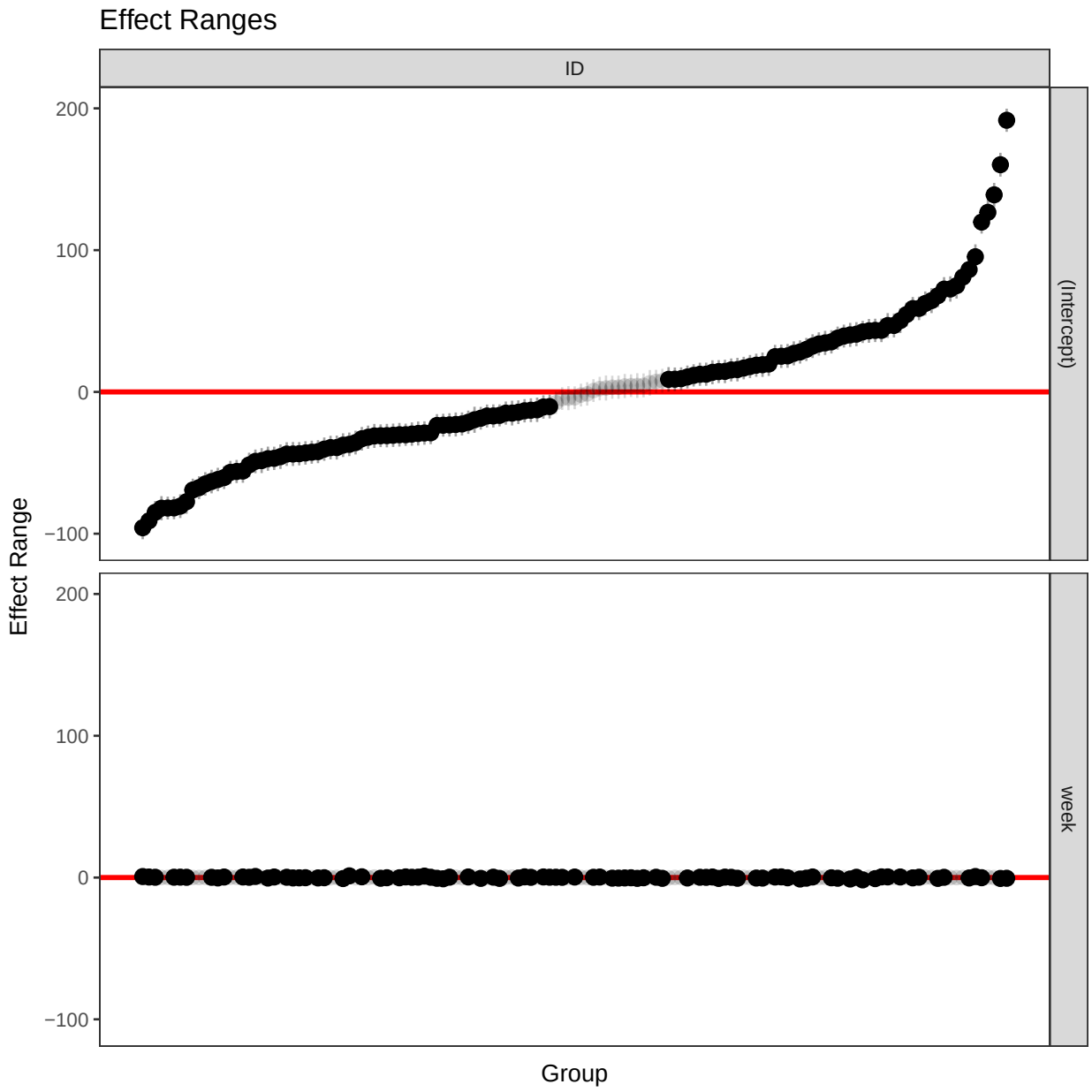

```
p1.gg2 <- re.sim %>% filter(term == "week")
ggplot(p1.gg2, aes(mean)) + geom_density()
```



7 Question 7

Create a catepilar plot of the random effects. Is there any person that seems odd in terms of a large standard errors around intercept and slope estimates?

```
p1 <- plotREsim(re.sim)
p1
```



Yes, there a couple that go from 100-200.

8 Question 8

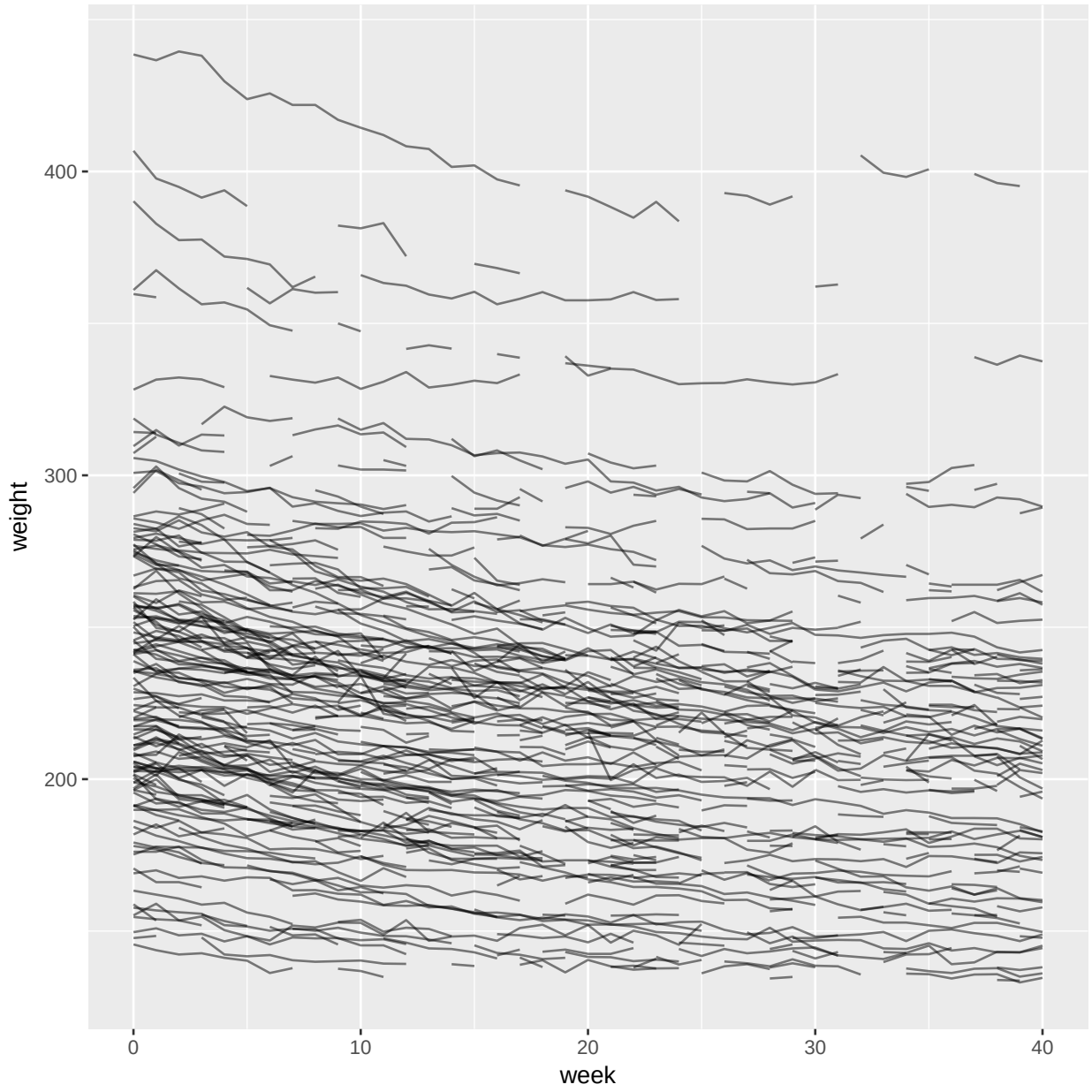
Create a plot of the trajectory, along with a spaghetti plot of each persons individual slope. Set the alpha level (transparency) on the individual slopes to make them easier to see.

```
library(tidyverse)

## Warning:  package 'tidyverse' was built under R version 3.3.2
```

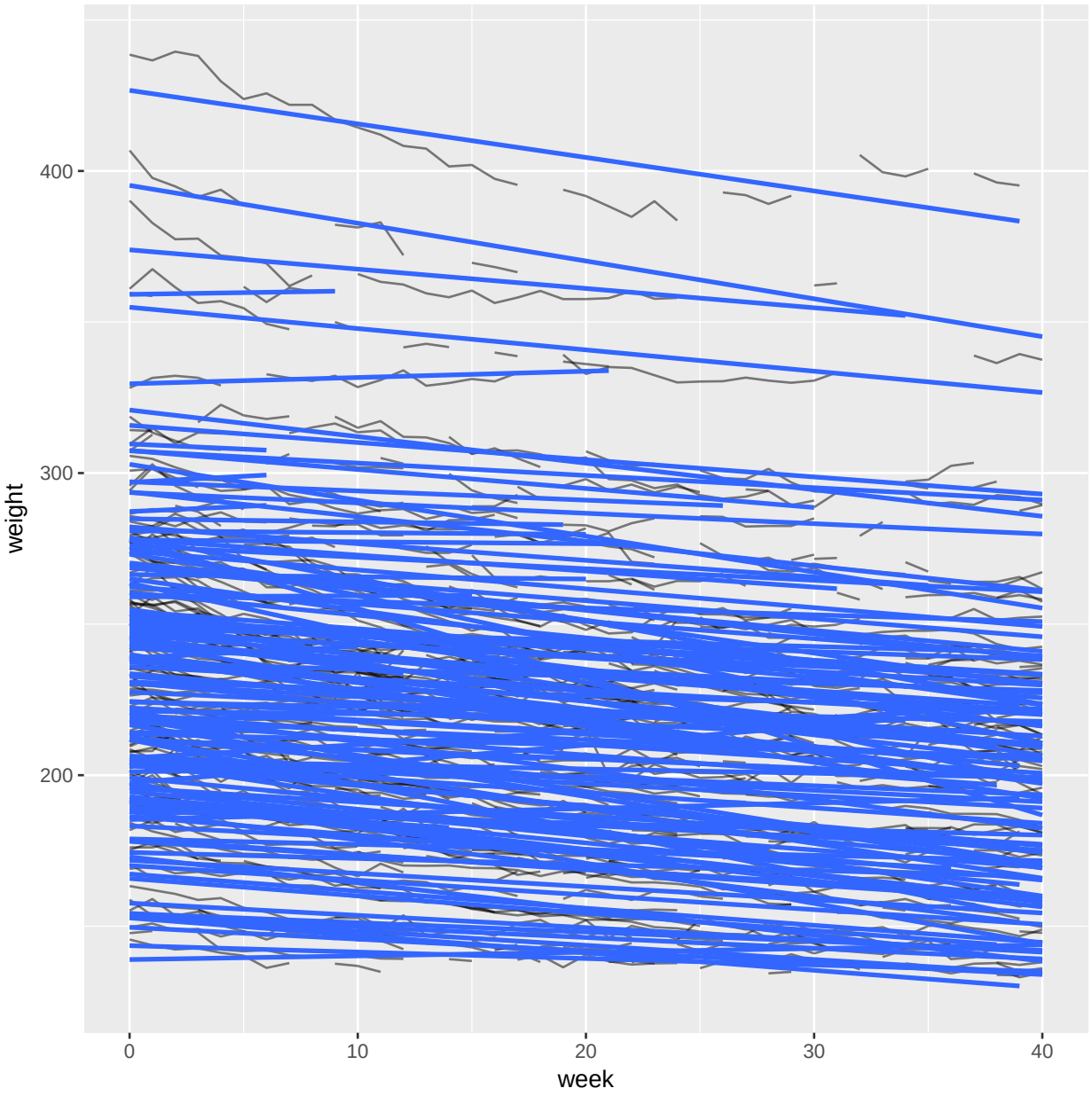
```
## Loading tidyverse:  tibble
## Loading tidyverse:  readr
## Loading tidyverse:  purrr
## Warning:  package 'tibble' was built under R version 3.3.2
## Warning:  package 'readr' was built under R version 3.3.2
## Warning:  package 'purrr' was built under R version 3.3.2
## Conflicts with tidy packages ----------------------------------------------
## arrange():    dplyr, plyr
## as.difftime():  lubridate, base
## compact():    purrr, plyr
## count():      dplyr, plyr
## date():        lubridate, base
## expand():      tidyr, Matrix
## failwith():    dplyr, plyr
## filter():      dplyr, stats
## here():        lubridate, plyr
## id():          dplyr, plyr
## intersect():  lubridate, base
## lag():        dplyr, stats
## mutate():      dplyr, plyr
## rename():      dplyr, plyr
## select():      dplyr, MASS
## setdiff():    lubridate, base
## summarise():  dplyr, plyr
## summarize():  dplyr, plyr
## union():      lubridate, base
```

```
gg1 <- ggplot(wgt, aes(x = week, y = weight, group = ID)) + geom_line(alpha = 0.5)
gg1
```

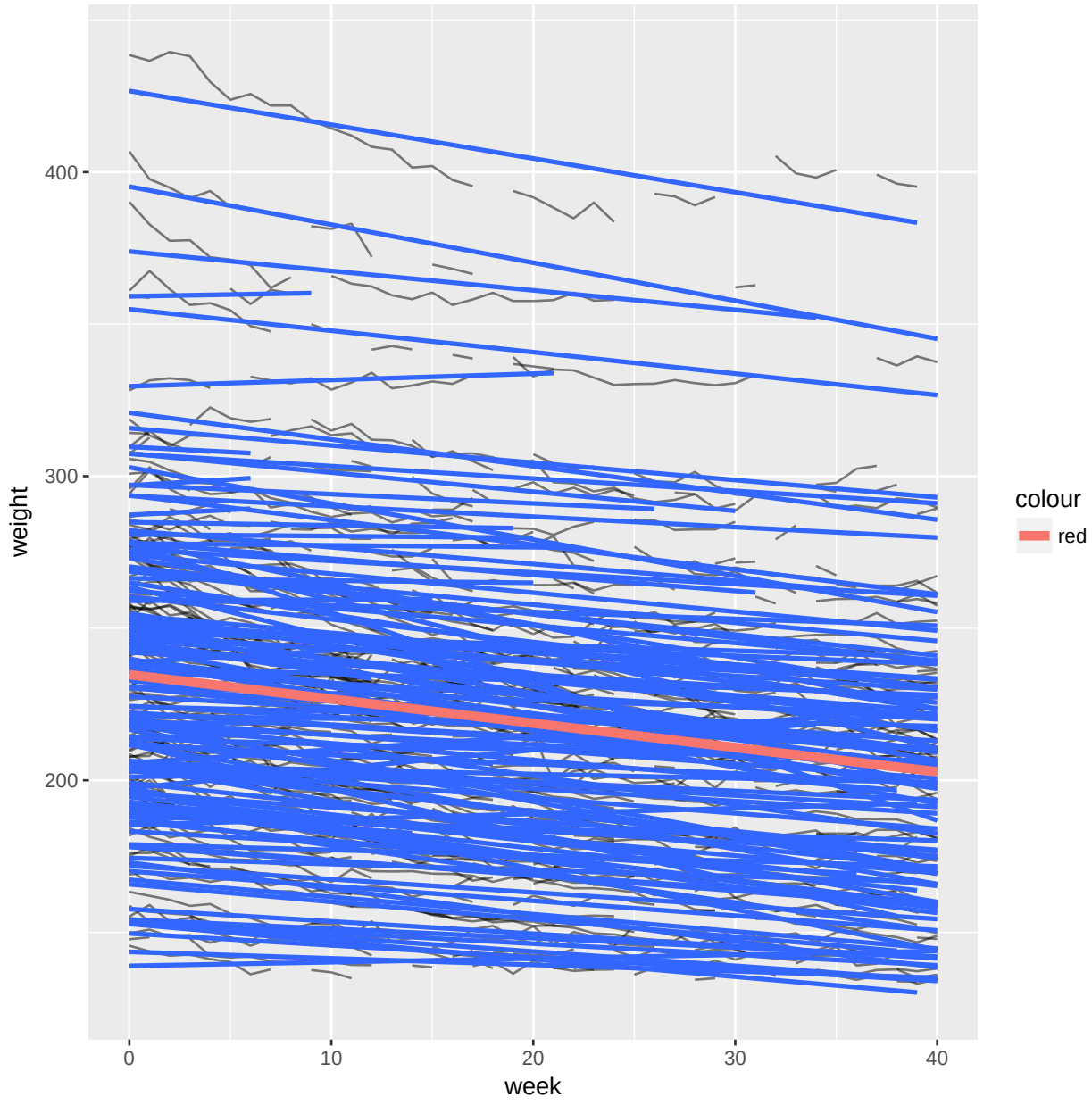
```
## Warning:  Removed 728 rows containing missing values (geom_path).
```




```
gg2 <- ggplot(wgt, aes(x = week, y = weight, group = ID)) + geom_line(alpha = 0.5) +
    stat_smooth(method = "lm", se = FALSE, alpha = 0.1)
gg2

## Warning:  Removed 1799 rows containing non-finite values (stat_smooth).
## Warning:  Removed 728 rows containing missing values (geom_path).
```




```
# can't get the smoothed trajectories to lighten

gg3 <- gg2 + stat_smooth(data = wgt, aes(x = week, y = weight, group = 1, colour = "red"),
    method = "lm", size = 2, se = FALSE)
print(gg3)

## Warning:  Removed 1799 rows containing non-finite values (stat_smooth).
## Warning:  Removed 1799 rows containing non-finite values (stat_smooth).
## Warning:  Removed 728 rows containing missing values (geom_path).
```


400
300
200
weight
0
10
20
30
40
week
colour
red