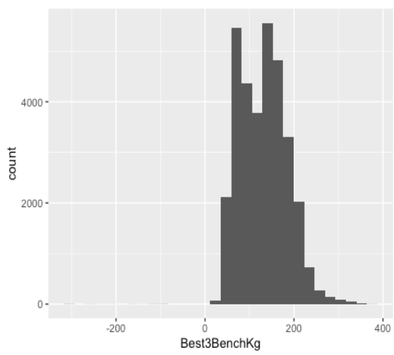
STAT-313 Oliver Fisher
Dr. Ramler 12/14/2023

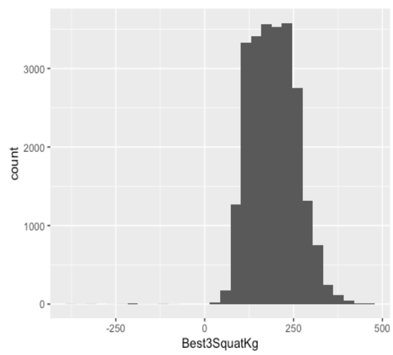
Final Project Report

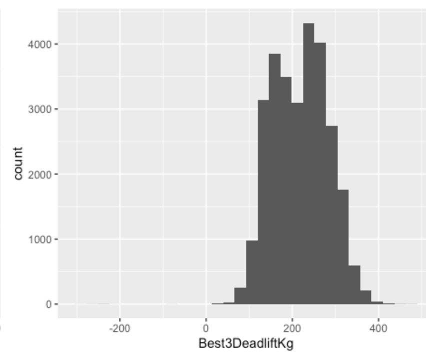
```r
library(tidyverse)

lift <- read_csv("data/powerlifting_313.csv")
```

Response Variables

- Best3BenchKg
- Best3SquatKg
- Best3DeadliftKg







-> All none Normal Distribution, potentially dependent on the individual lifter (Name)

Research Questions (Prelimiary)

- <u>Research Questions 1:</u> Will higher BodyweightKg, after accounting for Name, Equipment, Sex, and Age be associated with a higher Best3BenchKg?

-> Model 1: BodyweightKg at Level one with covariant indicator for single-ply Equipment type and indicator for raw Equipment type at Level One; and Age as Level Two predictor for intercept and all slope terms

- <u>Research Questions 2:</u> Will Will higher BodyweightKg, after accounting for Name, Equipment, Sex, and Age be associated with a higher Best3SquatKg?

-> Model 2: BodyweightKg at Level one with covariant indicator for single-ply Equipment type and indicator for raw Equipment type at Level One; Age as Level Two predictor for intercept and all slope terms

- <u>Research Questions 3:</u> Will Will higher BodyweightKg, after accounting for Name, Equipment, Sex, and Age be associated with a higher Best3DeadliftKg?

-> Model 3: BodyweightKg at Level one with covariant indicator for single-ply Equipment type and indicator for raw Equipment type at Level One; Age as Level Two predictor for intercept and all slope terms

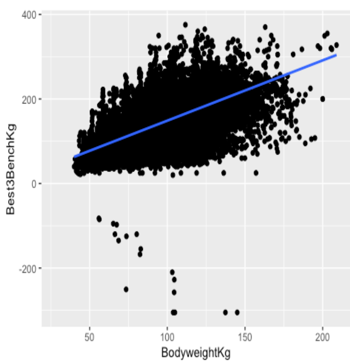
STAT-313
Dr. Ramler

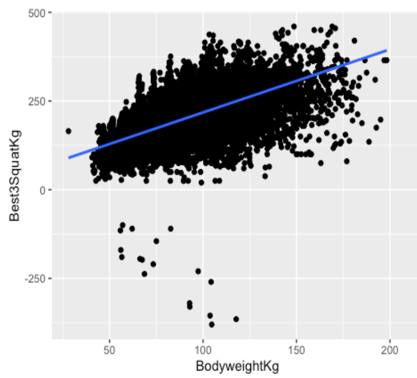
Oliver Fisher
12/14/2023

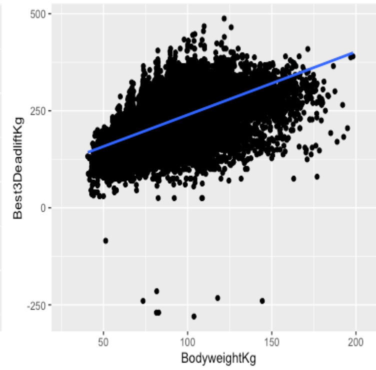
Why Multilevel Model Yes

The reason a multilevel models are appropriate is based on the nature of the data we wish to analyze, which is comprised of two levels. Level 1 which describes the individual performances ie. the lifts (deadlift, squat and, bench press) with covariate being equipment used during said lift and the recorded BodyweightKg of the lifter at the lift. Level 2 which pertains to the individual lifters and the level 2 covariates of their age and sex. In the model we wished to use level one and two covariate explain level 1 response variable (lift) while accounting for a level 2 variable between groups (Name).

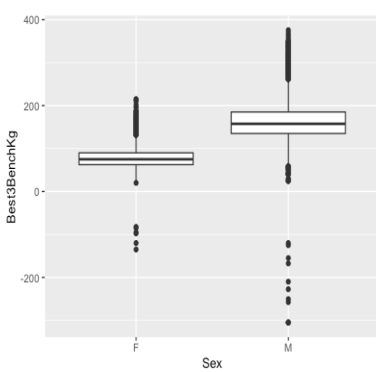
Exploratory Data Analysis

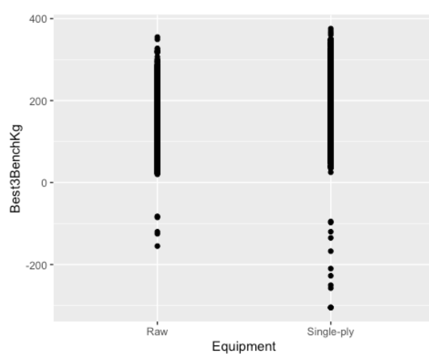


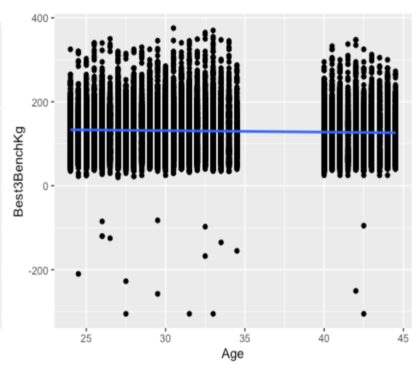




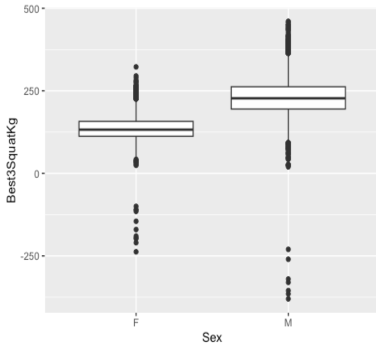
From the BodyweightKg Vs Best3BenchKg dot plot we can see a positive trend which might suggest an association. From the BodyweightKg Vs Best3SquatKg dot plot we can see a positive trend which might suggest an association. From the BodyweightKg Vs Best3DeadliftKg dot plot we can see a positive trend which might suggest an association.

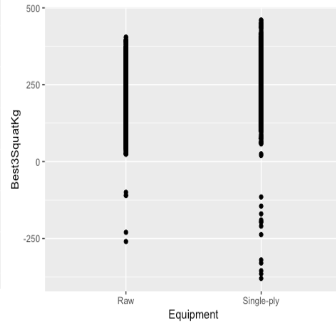


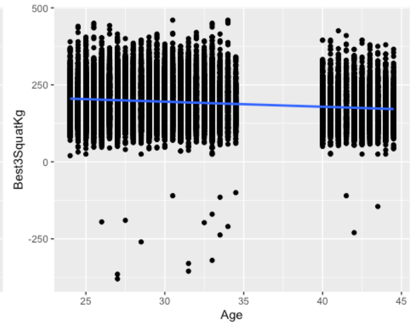




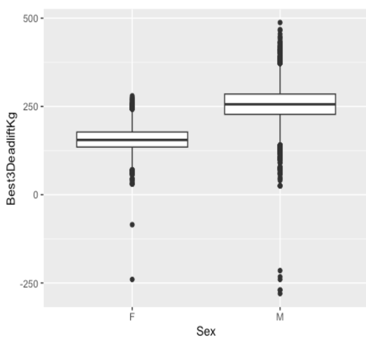
From the boxplot Sex vs Best3BenchKg we can see that there is a very clear distinction in weight lifted based on sex. From the plot Equipment vs Best3BenchKg the use of equipment appears to have a higher weight. From the plot Age vs Best3BenchKg there does not appear to me much significance between age and weight benched, might exclude age from the model.

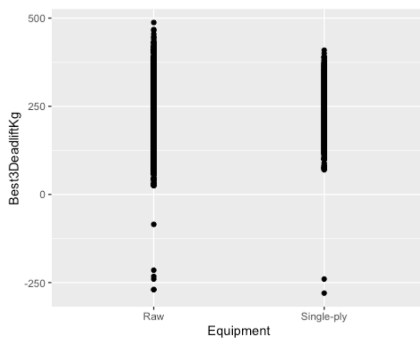


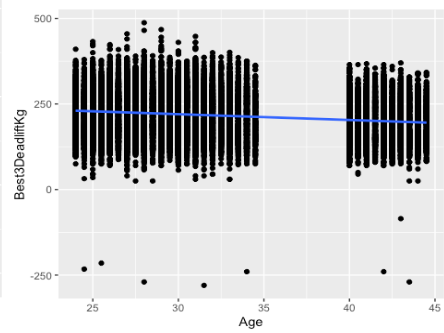




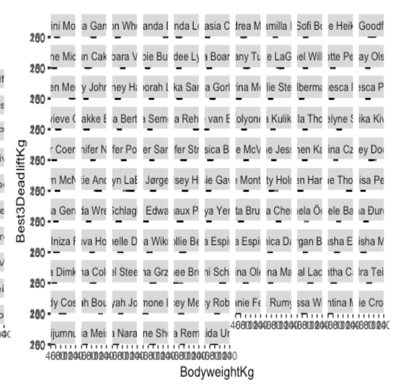
From the boxplot Sex vs Best3SquatKg we can see that there is a very clear distinction in weight lifted based on sex. From the plot Equipment vs Best3SquatKg the use of equipment appears to have a higher weight. From the plot Age vs Best3SquatKg there does appear to me a minor negative association between age and weight squatted.







From the boxplot Sex vs Best3DeadliftKg we can see that there is a very clear distinction in weight lifted based on sex. From the plot Equipment vs Best3DeadliftKg the use of equipment appears to have less weight the raw. From the plot Age vs Best3DeadliftKg there does appear to me a minor negative association between age and weight deadlifted



Given the sheer number of individuals, before we can evaluate the necessity of accommodating for individuals within the model we must first condense it to a group that can be faceted. To do so we have filtered the data to only include athletes who have competed the most amount of times (10) and only women. Those this significantly reduces the representation of the data we feel that individuals and their deviations can sufficiently be determined with any group of athletes. In this circumstance since our only necessity is to be able to see the plots compared against other athletes. Given the above plots we feel that it is necessary to use in multilevel model accounting for individuals given the major variation in BodyweightKg between individuals.

Unconditional Means Models

```
model_1 <- lmer(Best3BenchKg ~ 1 + (1 | Name), REML = T, data = lift)
summary(model_1)

model_2 <- lmer(Best3SquatKg ~ 1 + (1 | Name), REML = F, data = lift)
summary(model_2)

model_3 <- lmer(Best3DeadliftKg ~ 1 + (1 | Name), REML = F, data = lift)
summary(model_3)
```

The unconditional means models above all show that there is correlation for each individual

Fit Multilevel Model

```
model_1a <- lmer(Best3BenchKg ~ BodyweightKg  + (BodyweightKg | Name), REML =
F, data = lift)

model_2a <- lmer(Best3SquatKg ~ BodyweightKg + (BodyweightKg | Name), REML =
F, data = lift)

model_3a <- lmer(Best3DeadliftKg ~ BodyweightKg + (BodyweightKg | Name), REML
= F, data = lift)
```

Fit Multilevel Model Using Level Two Covariates

```
cat_lift <- lift |>
  mutate(male = ifelse(Sex == "M", 1, 0),
         female = ifelse(Sex == "F", 1, 0)) |>
  mutate(Raw = ifelse(Equipment == "Raw", 1, 0),
         Single_ply = ifelse(Equipment == "Single-ply", 1, 0))

model_1b <- lmer(Best3BenchKg ~ BodyweightKg + male + female + Raw +
Single_ply + (BodyweightKg | Name), REML = F, data = cat_lift)

model_2b <- lmer(Best3SquatKg ~ BodyweightKg + Age + male + female + Raw +
Single_ply + (BodyweightKg | Name), REML = F, data = cat_lift)

model_3b <- lmer(Best3DeadliftKg ~ BodyweightKg + Age + male + female + Raw +
Single_ply + (BodyweightKg | Name), REML = F, data = cat_lift)
```

Likelihood Ratio Tests (Nested)

<u>Best3BenchKg:</u>

$H_o = \alpha_1 = \alpha_2 = 0$

H_a = at least one of $\alpha_i \neq 0$

```
theta <- -2*(-143049.1) - (-2*(-139120.0))
df <- 8 - 6
pchisq(theta, df, lower.tail = FALSE)

[1] 0
```

p-value ~ 0 -> We reject the Null hypothesis full model is a best model over the base model

<u>Best3SquatKg:</u>

$H_o = \alpha_1 = \alpha_2 = \alpha_3 = 0$

H_a = at least one of $\alpha_i \neq 0$

```
theta <- -2*(-114257.5) - (-2*(-112415.9))
df <- 9 - 6
pchisq(theta, df, lower.tail = FALSE)

[1] 0
```

p-value ~ 0 -> We reject the Null hypothesis full model is a best model over the base model

<u>Best3DeadliftKg:</u>

$H_o = \alpha_1 = \alpha_2 = \alpha_3 = 0$

H_a = at least one of $\alpha_i \neq 0$

```
theta <- -2*(-130340.2) - (-2*(-128511.2))
df <- 9 - 6
pchisq(theta, df, lower.tail = FALSE)

[1] 0
```

p-value ~ 0 -> We reject the Null hypothesis full model is a best model over the base model

Final Composite Models

$Best3BenchKg_{ij} = \alpha_0 + \alpha_1(male) + \alpha_2(Raw) + \beta_0(BodyweightKg) + u_i + v_i$

$\widehat{\alpha 0}$ = 44.61145 mean weight of Squat on first lift for male lifting Raw at Age 0

$\widehat{\alpha 1}$ = 56.92621 mean weight of Squat for male lifter

$\widehat{\alpha 2}$ = -30.67064 mean weight of Squat lifting Raw

$\widehat{\beta 0}$ = 0.88131 the mean (slope) change in body weight of lifter

$\widehat{\sigma u}$ = 7.3644 standard deviation of squat weight between lifts

$\widehat{\sigma v}$ = 0.3001 standard deviation of lifter's body weight between lifts

$\widehat{\delta uv}$ = - 0.21 correlation in weight of squat and the lift's rate of change bodyweight

$Best3SquatKg_{ij} = \alpha_0 + \alpha_1(male) + \alpha_2(Raw) + \alpha_3(Age) + \beta_0(BodyweightKg) + u_i + v_i$

$\widehat{\alpha 0}$ = 75.57 mean weight of Squat on first lift for male lifting Raw at Age 0

$\widehat{\alpha 1}$ = 61.47383 mean weight of Squat for male lifter

$\widehat{\alpha 2}$ = -32.83 mean weight of Squat lifting Raw

$\widehat{\alpha 3}$ = 0.03115 mean weight of Squat at Age 0

$\widehat{\beta 0}$ = 1.24013 the mean (slope) change in body weight of lifter

$\widehat{\sigma u}$ = 17.0333 standard deviation of squat weight between lifts

$\widehat{\sigma v}$ = 0.4083 standard deviation of lifter's body weight between lifts

$\widehat{\delta uv}$ = - 0.26 correlation in weight of squat and the lift's rate of change bodyweight

$Best3DeadliftKg_{ij} = \alpha_0 + \alpha_1(male) + \alpha_2(Raw) + \alpha_3(Age) + \beta_0(BodyweightKg) + u_i + v_i$

$\widehat{\alpha 0}$ = 85.07195 mean weight of Squat on first lift for male lifting Raw at Age 0

$\widehat{\alpha 1}$ = 74.96013 mean weight of Squat for male lifter

$\widehat{\alpha 2}$ = -7.91818 mean weight of Squat lifting Raw

$\widehat{\alpha 3}$ = 0.29126 mean weight of Squat at Age 0

$\widehat{\beta 0}$ = 0.94961 the mean (slope) change in body weight of lifter

$\widehat{\sigma u}$ = 21.1087 standard deviation of squat weight between lifts

$\widehat{\sigma v}$ = 0.3843 standard deviation of lifter's body weight between lifts

$\widehat{\delta uv}$ = - 0.22 correlation in weight of squat and the lift's rate of change bodyweight

Summary of Findings

After observing the preliminary plots, we identified first that body weight in relationship to the respective lift disciplines must be evaluated and modeled in respect to the individual lifters. After evaluating the individual covariates we found that all were significant to the end models except for age in relation to bench press which appeared to have no relation. After conducting the likelihood ratio tests we were able to determine in all three models the full models were most appropriate to explain lifting weight output for bench press, squats, and deadlifts. Within the models themselves we also found that males lift more than women on average. That lifting raw is subtractive to the potential best lift as well as Age (except in the case of bench press with respect to Age). The model suggests that in all three cases the most ideal scenario involves being male, lifting with a single-ply, and in general the heavier the lifter the better.