

# Lab 11

04/02/25

## Preparation for Lab 11:

- Please review the materials provided in [https://rconnect.math.montana.edu/Data\\_Wrangling/#section-relational-data-with-dplyr](https://rconnect.math.montana.edu/Data_Wrangling/#section-relational-data-with-dplyr)
- Pay particular attention to the discussion of `left_join` and think about how we might use it to join your repeated measures of grip strength with the demographics data set.
- For the remaining work this semester, we need you to update the version of `catstats2` to version 0.3.0. Please run the following code once. It will ask you to update other packages and that should not be needed, so you can go to the console and choose option 3 or enter on an empty line.

```
#remotes::install_github("greenwood-stat/catstats2", force = T)
library(catstats2)
packageVersion("catstats2")

## [1] '0.3.0'
```

- If the previous chunk reports 0.3.0, you can comment out the `remotes::` line for the rest of your work on the lab.

## Part 1: Grip strength potential interaction:

```
GripStrengthMeasurements <- read_excel("GripStrengthMeasurementsS25.xlsx")
GS <- GripStrengthMeasurements %>% drop_na(GripStrength) #Mostly cleans out empty rows

library(mi)
r1 <- missing_data.frame(as.data.frame(GS))
table(r1@patterns)

##
## nothing      Arm
##      263      1

tally(Hand~Arm, data = GS)

##      Arm
## Hand 90 down Down up Up <NA>
```

```
##      D  44   24   20 27 17    0
##     ND 44   25   18 27 17    1

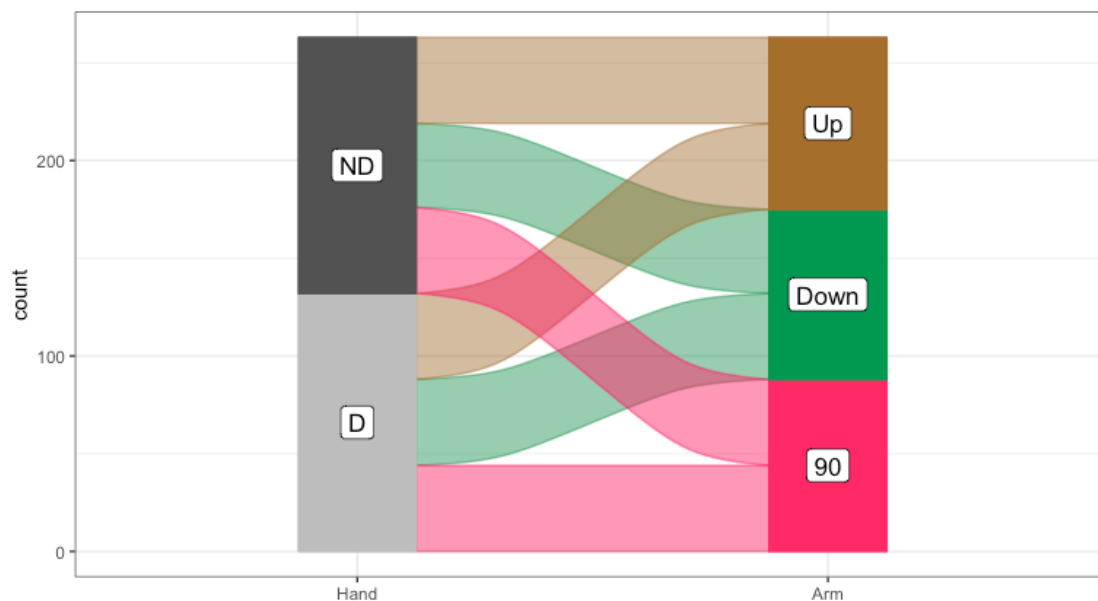
GS2 <- GS %>% drop_na(Arm)
GS2 <- GS2 %>% mutate(Arm = forcats::fct_collapse(Arm,
                                                    Up = c("Up", "up"),
                                                    Down = c("Down", "down"),
                                                    "90" = c("90", "90 degree")),
                      SubjectID = forcats::fct_collapse(SubjectID,
                                                         MDEar1 = c("MDEar1", "MDEar")),
                      Arm = factor(Arm),
                      Hand = factor(Hand),
                      OrderF = factor(Order))
)
tally(~Arm, data = GS2)

## Arm
##   90 Down  Up
##   88   87  88

tally(Arm ~ Hand, data = GS2)

##      Hand
## Arm    D ND
##   90   44 44
##  Down 44 43
##   Up   44 44

GS2 %>% dplyr::select(Hand, Arm) %>% alluvial_wide(bin_labels = "mean",
fill_by = "last_variable", bins = 6)
```

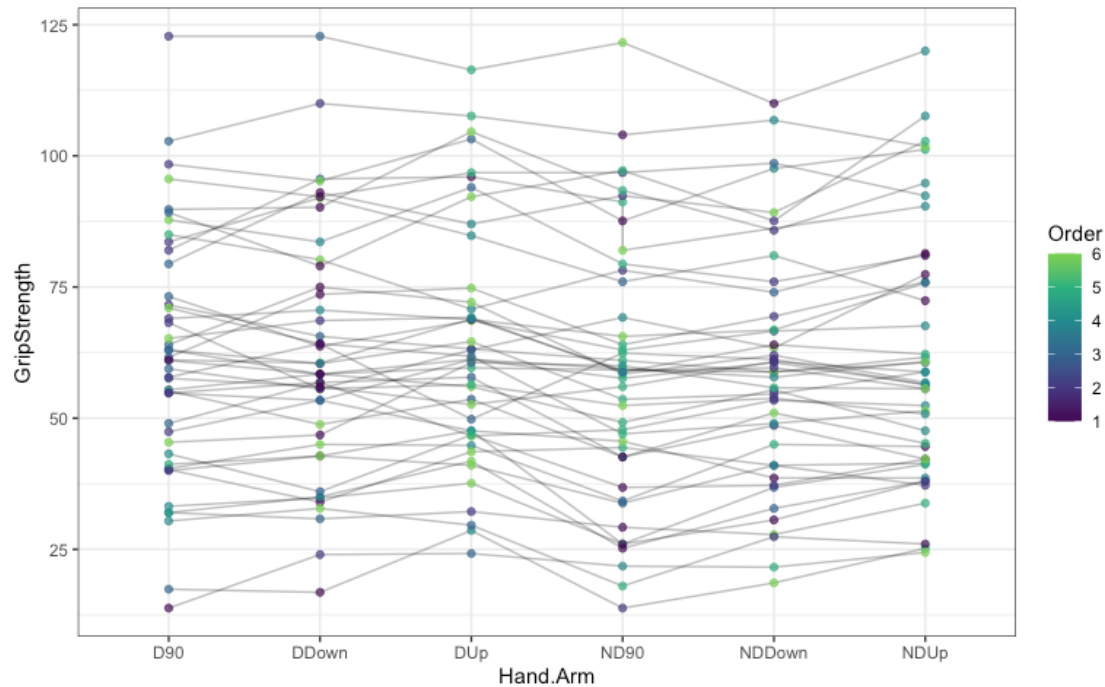


Number of flows: 6  
Original Dataframe reduced to 2.3 %  
Maximum weight of a single flow 16.7 %

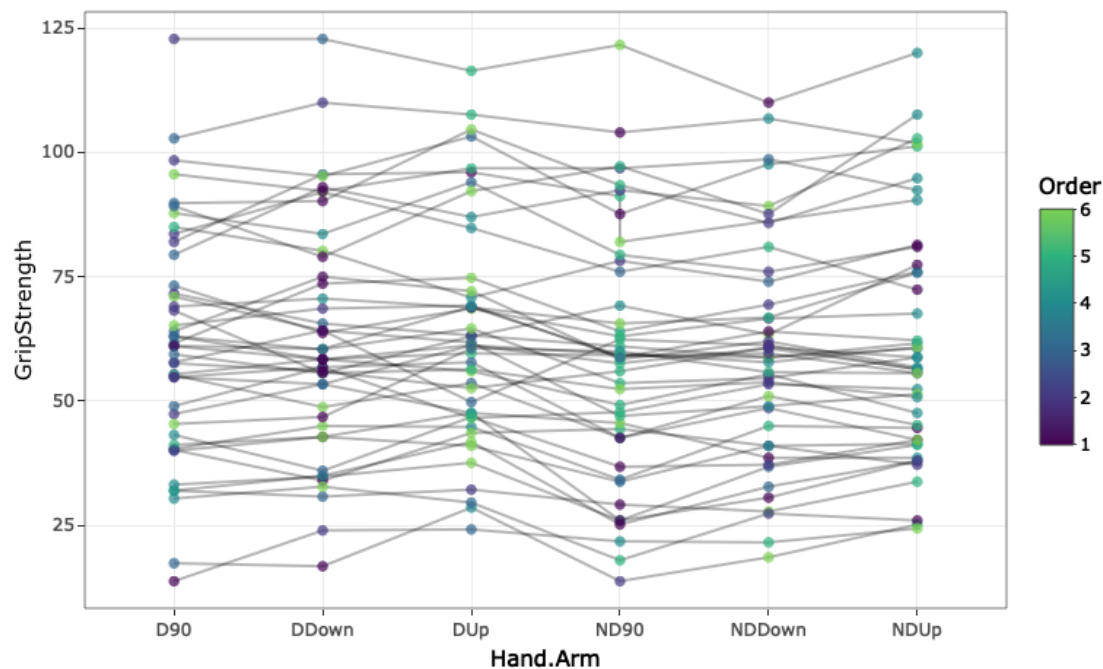
```
GS2 <- GS2 %>% mutate(Hand.Arm = factor(str_c(Hand, Arm)))
```

```
Spaghetti_GS2 <- GS2 %>% ggplot(mapping = aes(x = Hand.Arm, y = GripStrength,  
group = SubjectID)) +  
  geom_line(alpha = 0.3) +  
  geom_point(aes(color = Order), alpha = 0.7) +  
  scale_color_viridis_c(end = 0.8)
```

Spaghetti\_GS2



```
Spaghetti_GS2 %>% ggplotly()
```



### 1) What did the `fct_collapse` function do and why was it needed in the previous code?

'fct\_collapse' collapses factor levels into our own defined groups. This was needed in the code to make sure we are accounting for observations that spelled their inputs differently. For example, it defined inputs that put "Up" and "up" both as "Up". It did this for the subjectID for MDEarl when there was a misspelling as well.

**2) The addition of the line of code of `webshot::install_phantomjs(force = T)` should allow you to knit to Word even with interactive documents, but uncomment in the `ggplotly` lines if that is problematic. Find someone in your group or an observation that looks interesting and track that value across the treatment combinations. Does the subject tend to do better or worse on dominant or non-dominant hand across arm positions? Is there an optimal arm position for either or both hands for that subject? Now, look at the whole plot. What can you say about the differences in grip strength across hands (dominant/non-dominant), arm positions, and subjects?**

We are following OrlandoChamp! For this subject the highest gripstrength was observed when dominant hand was in the Up position. Other than this difference, the other combinations of dominant and nondominant in various positions were within 3lbs of each other. When looking at the whole plot, nondominant observations seem to be lower than dominant observations across subjects. There also seems to be a trend where 90degree observations have lower GripStrengths than down or up. This holds for both dominant and nondominant hands.

### 3) Is this a balanced design relative to the Hand and Arm variables?

No! (Technically) There are 44 observations for all combinations of hand and arm except for nondominant down which only has 43.

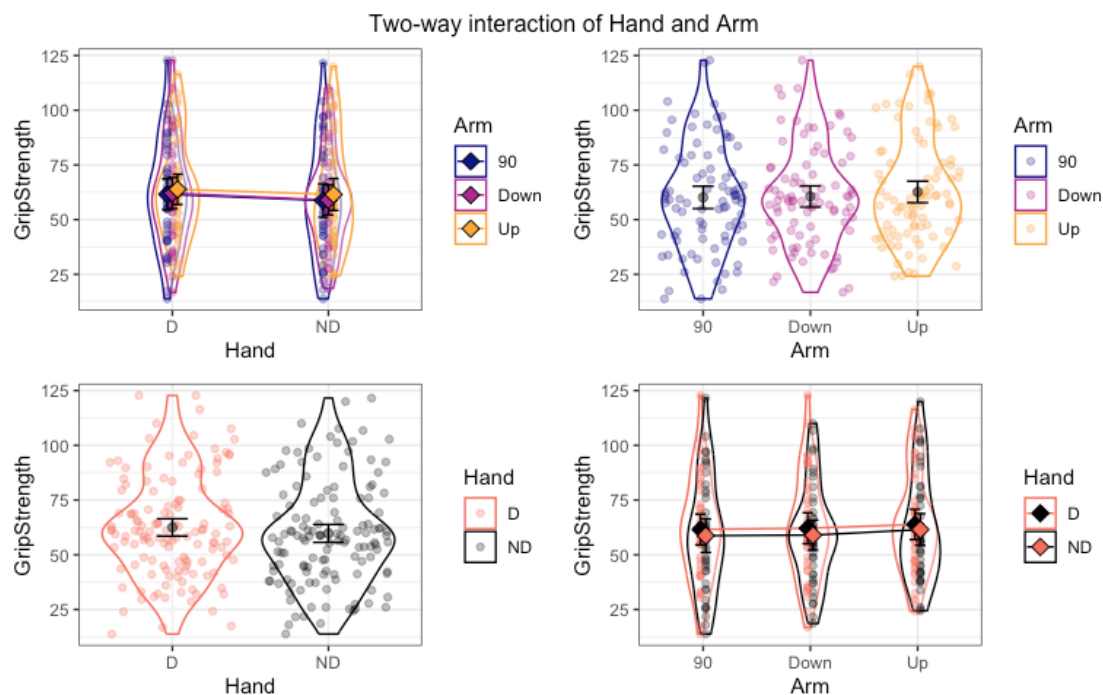
```
tally(Arm ~ Hand, data = GS2)
```

```
##      Hand
## Arm    D ND
##  90    44 44
## Down   44 43
## Up     44 44
```

**4) Make an interaction plot and discuss the potential for an interaction of Arm and Hand on the Grip Strength responses. If you think there is a potential interaction, discuss the pattern. If you don't think there is a potential interaction, discuss the possible pattern of results in the two enhanced stripcharts that provide a hint about the results for each of the main effects.**

Based on the interaction plot there does not seem to be a huge potential for an interaction between Arm and Hand. We say this because the lines in the upper left and lower right plots seem to be parallel to each other with not much difference in slope. Looking at the upper right and lower left stripcharts, we can see some patterns for the two main effects. In these it looks like the Up position had the highest observations and 90 degree lowest. Non-dominant hands also showed a general lower trend than the dominant hands.

```
ggintplot(response = "GripStrength", groupvars = c("Hand", "Arm"), data = GS2)
```



5) We will dive more deeply into linear mixed models in the coming weeks, but the following code uses the `lmer` function from the `lme4` package (with some modifications from the `lmerTest` package so you can see p-values) to fit a mixed model that incorporates subject as a “random” effect. You can interpret the results for the “fixed” effect (starting with arm and hand and their interaction on grip strength) as “conditional on” or “adjusted for” subject much like we did if it was included as a fixed effect. Report an evidence sentence for the interaction test from the following results and what it suggests for the interaction in the model.

There is weak evidence against the null hypothesis of no interaction between Hand and Arm on GripStrength ( $F_{2,214.01} = 0.377$ , p-value = 0.686), adjusting for subjects, so we would conclude there is not an interaction and drop it from the model.

```
lmer1 <- lmer(GripStrength ~ Hand*Arm + (1|SubjectID), data = GS2)

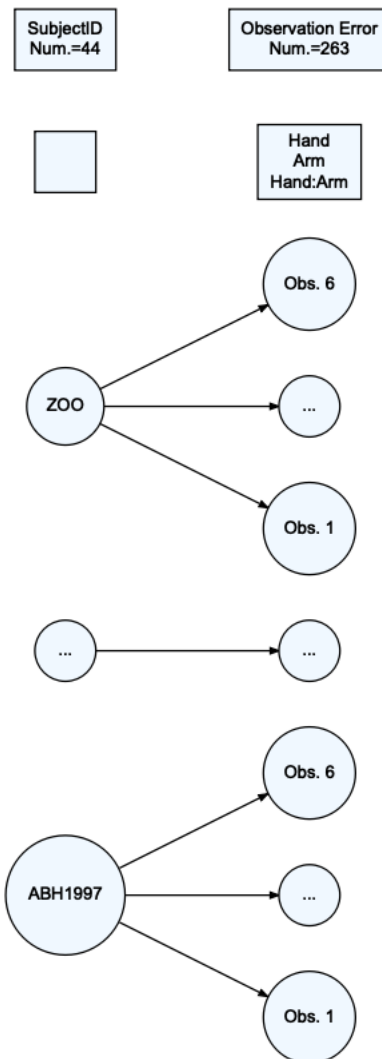
Anova(lmer1, test.statistic = "F")

## Analysis of Deviance Table (Type II Wald F tests with Kenward-Roger df)
##
## Response: GripStrength
##              F Df Df.res    Pr(>F)
## Hand       19.928  1 214.00  1.3e-05
## Arm         7.275  2 214.01 0.0008776
## Hand:Arm    0.377  2 214.01 0.6863821
```

6) Run the following code to generate a `model_diagram` for this model. How can you explain the location for the Hand, Arm, and Hand:Arm interaction in the plot?

Hand, Arm, and Hand:Arm are located at the top of the right column of the model diagram. This is because these are indicated at the observation level, not the subject level. The 6 observations per individual are a combination of hand and arm. If they were over the left column it would indicate that the subjects were split into arm and hand combinations. For example subject 1 would only do dominantUp and subject 2 would do nondominantDown, etc.

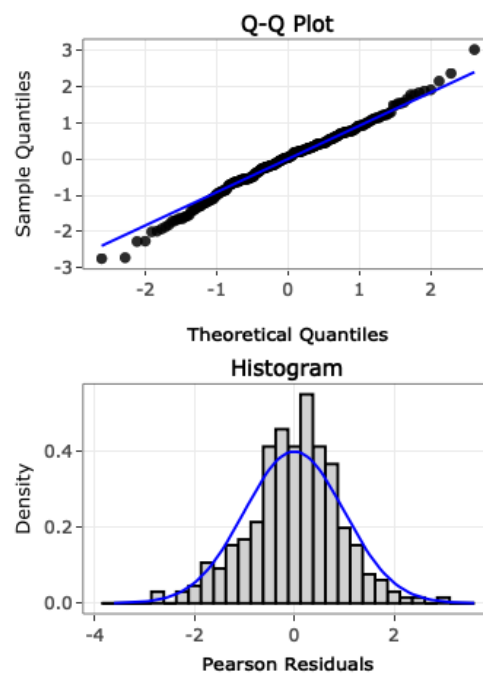
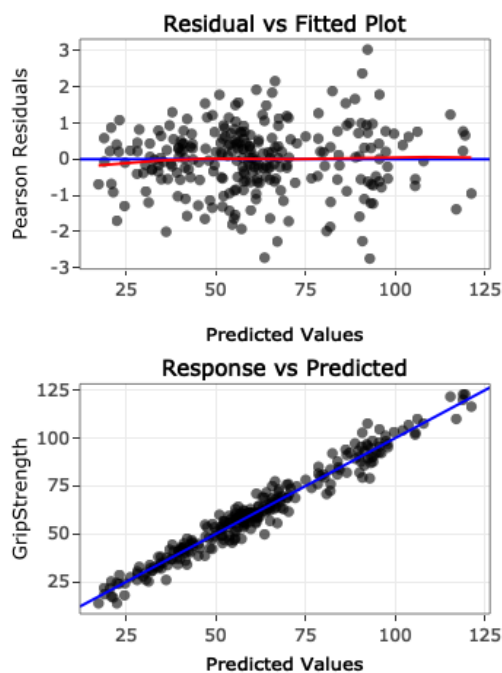
```
model_diagram(lmer1, heightVal = 800)
```



**7) An interactive residual panel is available below. Which subject(s) had the largest and smallest observed residuals (so were most above/below what the model predicted for them)? Generally how large/small was the fitted value for the two most extreme residuals?**

The subject BostonWolverine had both the highest and lowest observed residuals. This individual had an observation with residuals of both 3.027 and -2.74. The fitted values for these points were 92.27 and 92.88, respectively. The model here predicted values that were generally 3-7 lbs different.

```
resid_interact(lmer1, c("resid", "qq", "yvp", "hist"))
```



## Part II: Merging (left\_join-ing) the demographics data

The following code will read in and (eventually) join the demographics data.

```
dim(GS2)
## [1] 263 7

Demographics <- read_excel("DemographicsS25.xlsx")
Demographics <- Demographics %>% dplyr::select(-c(12:13)) #Because of a
"note" left in column 13...
dim(Demographics)
## [1] 44 11
```

**8) The following code highlights a potential issue with one of the subject IDs. Fix the problem in the Demographics data.frame using R code.**

```
data.frame(sort(unique(GS2$SubjectID)), sort(unique(Demographics$SubjectID)))

##      sort.unique.GS2.SubjectID.. sort.unique.Demographics.SubjectID..
## 1                ABH1997                ABH1997
## 2                AFEDU2020                AFEDU2020
## 3                AustinLions                AustinLions
## 4                AZBlackhawks                AZBlackhawks
## 5                  AZRams                  AZRams
## 6                BelknapBull                BelknapBull
## 7                BostonLakers                BostonLakers
## 8                BostonWolverines            BostonWolverines
## 9                  CAJustinBieber            CAJustinBieber
```



```

## 10      CenntenialWarriors      CentennialWarriors
## 11      COMichealFranti        COMichealFranti
## 12      DCSNAP                 DCSNAP
## 13      DenverCoyotes         DenverCoyotes
## 14      DenverGalaxy          DenverGalaxy
## 15      FLTaller               FLTaller
## 16      GASNAP                 GASNAP
## 17      HYENA                  HYENA
## 18      IAGarthBrooks         IAGarthBrooks
## 19      IDAJR                 IDAJR
## 20      IDPAD                  IDPAD
## 21      LACROIX               LACROIX
## 22      LADeathGrips          LADeath
## 23      MDEarl                MDEarl
## 24      NDLawrence            NDLawrence
## 25      NewYorkLaezeb         NewYorkLaezeb
## 26      NMSnuggies            NMSnuggies
## 27      NVKP                  NVKP
## 28      OrlandoChamp         OrlandoChamp
## 29      ORMichelangelo        ORMichelangelo
## 30      PATaller              PATaller
## 31      SanDiegoBobcat        SanDiegoBobcat
## 32      UTSNAP                 UTSNAP
## 33      VancouverMcLaren      VancouverMcLaren
## 34      VTPink                VTPink
## 35      WATaller              WATaller
## 36      WinonaBadgers         WinonaBadgers
## 37      WIWiggles             WIWiggles
## 38      WYBeavers             WYBeavers
## 39      WYIW                  WYIW
## 40      WYJourney             WYJourney
## 41      WYSlipKnot            WYSlipKnot
## 42      WYSNAP                WYSNAP
## 43      WYTaller              WYTaller
## 44      ZOO                   ZOO

which(sort(unique(GS2$SubjectID)) != sort(unique(Demographics$SubjectID)))

## [1] 10 22

# Fix the problem in Demographics to match GS2:
Demographics <- Demographics %>% mutate(SubjectID =
  forcats::fct_recode(SubjectID,
    LADeathGrips = "LADeath"),
  SubjectID =
    forcats::fct_recode(SubjectID,
      CenntenialWarriors =
        "CentennialWarriors"))

```

*#Check that problem was fixed:*

```
data.frame(sort(unique(GS2$SubjectID)), sort(unique(Demographics$SubjectID)))
```

```
##      sort.unique.GS2.SubjectID.. sort.unique.Demographics.SubjectID..  
## 1              ABH1997              ABH1997  
## 2              AFEDU2020              AFEDU2020  
## 3              AustinLions              AustinLions  
## 4              AZBlackhawks              AZBlackhawks  
## 5              AZRams              AZRams  
## 6              BelknapBull              BelknapBull  
## 7              BostonLakers              BostonLakers  
## 8              BostonWolverines              BostonWolverines  
## 9              CAJustinBieber              CAJustinBieber  
## 10             CenntennialWarriors              CentennialWarriors  
## 11             COMichealFranti              COMichealFranti  
## 12             DCSNAP              DCSNAP  
## 13             DenverCoyotes              DenverCoyotes  
## 14             DenverGalaxy              DenverGalaxy  
## 15             FLTaller              FLTaller  
## 16             GASNAP              GASNAP  
## 17             HYENA              HYENA  
## 18             IAGarthBrooks              IAGarthBrooks  
## 19             IDAJR              IDAJR  
## 20             IDPAD              IDPAD  
## 21             LACROIX              LACROIX  
## 22             LADeathGrips              LADeathGrips  
## 23             MDEarl              MDEarl  
## 24             NDLawrence              NDLawrence  
## 25             NewYorkLaeeb              NewYorkLaeeb  
## 26             NMSnuggies              NMSnuggies  
## 27             NVKP              NVKP  
## 28             OrlandoChamp              OrlandoChamp  
## 29             ORMichelangelo              ORMichelangelo  
## 30             PATaller              PATaller  
## 31             SanDiegoBobcat              SanDiegoBobcat  
## 32             UTSNAP              UTSNAP  
## 33             VancouverMcLaren              VancouverMcLaren  
## 34             VTPink              VTPink  
## 35             WATaller              WATaller  
## 36             WinonaBadgers              WinonaBadgers  
## 37             WIWiggles              WIWiggles  
## 38             WYBeavers              WYBeavers  
## 39             WYIW              WYIW  
## 40             WYJourney              WYJourney  
## 41             WYSlipKnot              WYSlipKnot  
## 42             WYSNAP              WYSNAP  
## 43             WYTaller              WYTaller  
## 44             ZOO              ZOO
```

```

Demographics <- Demographics %>% mutate(Weights = factor(Weights))

Demographics <- Demographics %>% mutate(Weights = fct_collapse(Weights,
no = c("no", "No"),
      yes = c("yes", "Yes")))

#Join Demographics to the GS2 repeated measures data
combined <- left_join(x = GS2, y = Demographics, by = "SubjectID")

combinedR <- combined %>% drop_na()

combinedR <- combinedR %>% mutate(forearm_bins =
      factor(cut_number(ForearmLength, n = 3)),
      epworth_bins =
      factor(cut_number(Epworth, n = 3)),
      balance_bins =
      factor(cut_number(BalanceTime, n = 3)))

dim(combinedR)

## [1] 251 20

tally(~SubjectID, data = combinedR)

## SubjectID
##          ABH1997          AFEDU2020          AustinLions
AZBlackhawks
##              6              6              6
6
##          AZRams          BelknapBull          BostonLakers
BostonWolverines
##              6              6              6
6
##      CAJustinBieber CenntenialWarriors      COMichealFranti
DCSNAP
##              6              0              5
6
##      DenverCoyotes          DenverGalaxy          FLTaller
GASNAP
##              6              6              6
6
##          HYENA          IAGarthBrooks          IDAJR
IDPAD
##              6              6              6
6
##          LACROIX          LADeathGrips          MDEarl
NDLawrence
##              6              6              0
6
##      NewYorkLaeeb          NMSnuggies          NVKP
OrlandoChamp

```

```
##          6          6          6
6
##   ORMichelangelo   PATaller   SanDiegoBobcat
UTSNAP
##          6          6          6
6
##   VancouverMcLaren   VTPink   WATaller
WinonaBadgers
##          6          6          6
6
##          WIWiggles   WYBeavers   WYIW
WYJourney
##          6          6          6
6
##          WYSlipKnot   WYSNAP   WYTaller
ZOO
##          6          6          6
6
## CentennialWarriors
##          0
```

**9) What is the sample size before and after the left\_join? How many subjects are in the data set as analyzed in combinedR?**

```
unique(combinedR$SubjectID)
```

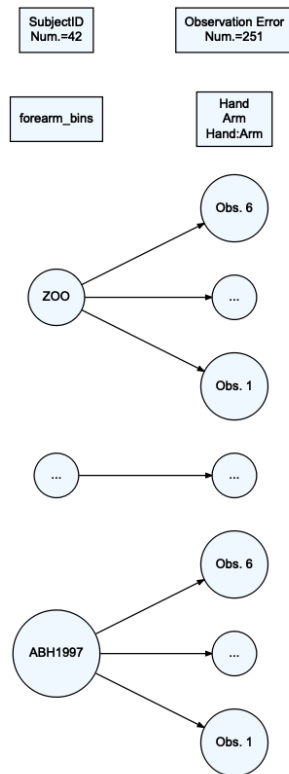
```
## [1] GASNAP          WYSNAP          UTSNAP          DCSNAP
## [5] COMichealFranti  LADeathGrips    NMSnuggies      CAJustinBieber
## [9] ORMichelangelo   VTPink          WYSlipKnot      IAGarthBrooks
## [13] WYJourney        WIWiggles       NDLawrence      WATaller
## [17] FLTaller         PATaller        WYTaller        NVKP
## [21] IDAJR           WYIW           IDPAD           LACROIX
## [25] VancouverMcLaren ZOO            HYENA           AFEDU2020
## [29] ABH1997          BelknapBull     DenverGalaxy     OrlandoChamp
## [33] AustinLions      BostonWolverines WinonaBadgers    NewYorkLaeeb
## [37] DenverCoyotes    SanDiegoBobcat  AZRams          BostonLakers
## [41] AZBlackhawks     WYBeavers
## 45 Levels: ABH1997 AFEDU2020 AustinLions AZBlackhawks AZRams ...
CentennialWarriors
```

**10) Now we can incorporate forearm binned variables into the model using the forearm\_bins, which we will treat as a fixed effect. Make a model\_diagram from the provided model and explain/discuss the location of the fixed effects in it.**

```
combinedR <- combinedR %>% mutate(
  forearm_bins = fct_recode(forearm_bins,
    low = "[9.25,10]",
    medium = "(10,11]",
    high = "(11,13.1]")
)
```

```
lmer2 <- lmer(GripStrength ~ Hand*Arm + forearm_bins + (1|SubjectID), data =
combinedR)
```

```
model_diagram(lmer2, heightVal = 800)
```



11) The following provides the theoretical and estimated fixed effect part of the model, except does not add the subscripts for the fixed effects or define the distributions of the random effects (see all locations with ?). Add the definition of the random effects and subscripts in the appropriate places based on the previous model diagram.

```
lmer2 %>% tbl_regression(intercept = T)
```

Characteristic	Beta	95% CI <sup>1</sup>	p-value
(Intercept)	47	38, 56	<0.001
Hand			
D	—	—	
ND	-3.5	-5.7, -1.3	0.002
Arm			
90	—	—	
Down	0.19	-2.0, 2.4	0.9
Up	1.9	-0.30, 4.0	0.090
forearm_bins			
low	—	—	
medium	13	-0.49, 27	0.058
high	35	21, 48	<0.001
Hand * Arm			
ND * Down	1.2	-1.9, 4.3	0.4
ND * Up	1.3	-1.7, 4.4	0.4

<sup>1</sup>CI = Confidence Interval

- Theoretical model:  $GripStrength_{ij} = \mu_{ij} + Subject_i + \epsilon_{ij}$ 
  - where  $Subject_i \sim ?$  and  $\epsilon_{ij} \sim ?$  and  $i = 1, \dots, I$  subjects and  $j = 1, \dots, n_i$  for the  $j^{th}$  observation on the  $i^{th}$  subject.
- Estimated model for  $\mu_{ij}$ :
  - Result:  $\hat{\mu}_{ij} = 47 - 3.7I_{Hand=ND,?} + 0.25I_{Arm=Down,?} + 2.0I_{Arm=Up,?} + 12I_{forearmbins=medium,?} + 35I_{forearmbins=high,?} + 1.4I_{Hand=ND,?}I_{Arm=Down,?} + 1.5I_{Hand=ND,?}I_{Arm=Up,?}$

**12) Note any additional resources used to complete this lab or NONE.**