power_2

Elias

May 10, 2018

Power project

Here I'll analyse some data from the open powerlifting project.

- Downloaded data from https://www.openpowerlifting.org/data.html on May 9th, 2018
- Let's play

```
library('lme4')
```

```
## Loading required package: Matrix
```

```
library(tidyr)
```

```
##
## Attaching package: 'tidyr'
```

```
## The following object is masked from 'package:Matrix':
##
##     expand
```

```
library(RColorBrewer)
library(ggplot2)
library(dplyr)
```

```
##
## 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(gtools)
library(tinytex)

open<-read.csv("~/power/data/open/openpowerlifting/openpowerlifting2.csv",head=TRUE,stringsAsFactors = F
str(open)
```

```
## 'data.frame':    579174 obs. of  30 variables:
##  $ MeetID        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ LifterID      : int  1 2 2 2 3 4 5 5 6 6 ...
##  $ Name          : chr  "Angie Belk Terry" "Dawn Bogart" "Dawn Bogart" "Dawn Bogart" ...
##  $ Sex           : chr  "F" "F" "F" "F" ...
##  $ Event         : chr  "SBD" "SBD" "SBD" "B" ...
##  $ Equipment     : chr  "Wraps" "Single-ply" "Single-ply" "Raw" ...
##  $ Age           : num  47 42 42 42 18 28 60 60 52 52 ...
##  $ Division      : chr  "Mst 45-49" "Mst 40-44" "Open Senior" "Open Senior" ...
```

```
##  $ BodyweightKg  : num  59.6 58.5 58.5 58.5 63.7 ...
##  $ WeightClassKg : chr  "60" "60" "60" "60" ...
##  $ Squat1Kg      : num  38.6 120.2 120.2 NA NA ...
##  $ Squat2Kg      : num  47.6 136.1 136.1 NA NA ...
##  $ Squat3Kg      : num  -54.4 142.9 142.9 NA NA ...
##  $ Squat4Kg      : num  NA NA NA NA NA ...
##  $ BestSquatKg   : num  47.6 142.9 142.9 NA NA ...
##  $ Bench1Kg      : num  15.9 88.5 88.5 88.5 29.5 ...
##  $ Bench2Kg      : num  20.4 95.2 95.2 95.2 31.8 ...
##  $ Bench3Kg      : num  -24.9 -97.5 -97.5 -97.5 -34 ...
##  $ Bench4Kg      : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ BestBenchKg   : num  20.4 95.2 95.2 95.2 31.8 ...
##  $ Deadlift1Kg   : num  61.2 136.1 136.1 NA 90.7 ...
##  $ Deadlift2Kg   : num  70.3 149.7 149.7 NA -97.5 ...
##  $ Deadlift3Kg   : num  -77.1 163.3 163.3 NA NA ...
##  $ Deadlift4Kg   : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ BestDeadliftKg: num  70.3 163.3 163.3 NA 90.7 ...
##  $ TotalKg       : num  138.3 401.4 401.4 95.2 122.5 ...
##  $ Place         : chr  "1" "1" "1" "1" ...
##  $ Wilks         : num  155 456 456 108 130 ...
##  $ McCulloch     : num  168 466 466 110 138 ...
##  $ Group         : chr  "47-49" "41-43" "41-43" "41-43" ...
```

- Let's start taking a look at the data.
  - First I will group people in age groups (3 year groups)
  - Then subset it into men and women

```
#assigning age groups
# group<-seq(from=range(open$Age,na.rm=TRUE)[[1]],
#            to=range(open$Age,na.rm=TRUE)[[2]],
#            by=3)
#
# group_range<-c()
# group2<-c(group-1,100)
# for(i in 1:length(group)){
#   if(i==length(group)){group_range[i]<-paste(c(group[i],"100"),collapse="-")
#   next()
#   }
#   group_range[i]<-(paste(c(group[i],group2[i+1]),collapse="-"))
# } #making these into ranges rather than just the minimums
#
# open<-cbind(open,Group=NA) #run once
# #adding a column of ranges
# for(i in 1:length(open$Age)){
#   if(is.na(open$Age[i])){next()}
#   for(j in 1:length(group)){
#     if(open$Age[i]>=group[j]&&open$Age[i]<=group2[j+1]){
#     open$Group[i]<-group_range[j]
#     }
#   }
# }

#writing table for future use because the chunk above takes FOREVER
#write.table(open,"~/power/data/open/openpowerlifting/openpowerlifting2.csv",quote=FALSE,row.names = FA
```

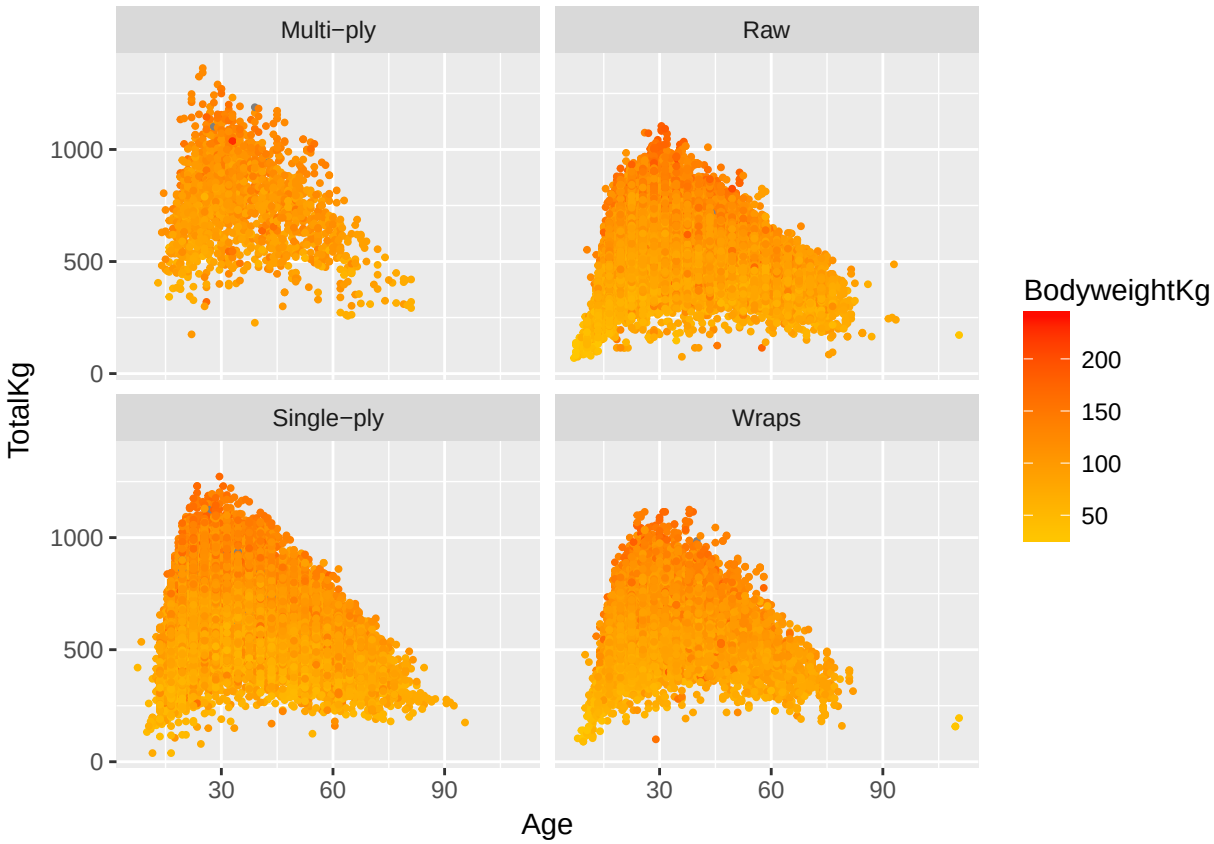
```
ord<-mixedorder(open$Group)
open2<-open[ord,]

open2$Group<-factor(open2$Group,levels=unique(open2$Group))

men<-open2 %>%
  filter(BestBenchKg>0 & BestDeadliftKg>0 & BestSquatKg>0) %>%
  filter(Sex=="M") %>%
  select(Equipment,Age,BodyweightKg,BestSquatKg,BestBenchKg,BestDeadliftKg,TotalKg,Wilks,Group)

women<-open2 %>%
  filter(BestBenchKg>0 & BestDeadliftKg>0 & BestSquatKg>0) %>%
  filter(Sex=="F") %>%
  select(Equipment,Age,BodyweightKg,BestSquatKg,BestBenchKg,BestDeadliftKg,TotalKg,Wilks,Group)
```

Age

- Now I'll try to figure out if Age has an effect on performance. . . and what that effect is
- I am starting with male data and then verify trends on females
- Purely driven by the fact that there is much more data for men, so the relationships can be seen more clearly

```
ggplot(men,
       aes(x=Age,y=TotalKg,color=BodyweightKg)) +
  geom_point(cex=.7) +
  facet_wrap(~Equipment) +
  scale_color_gradient2(guide="colourbar",low="yellow",mid="gold",high="red")
```

```
## Warning: Removed 176344 rows containing missing values (geom_point).
```



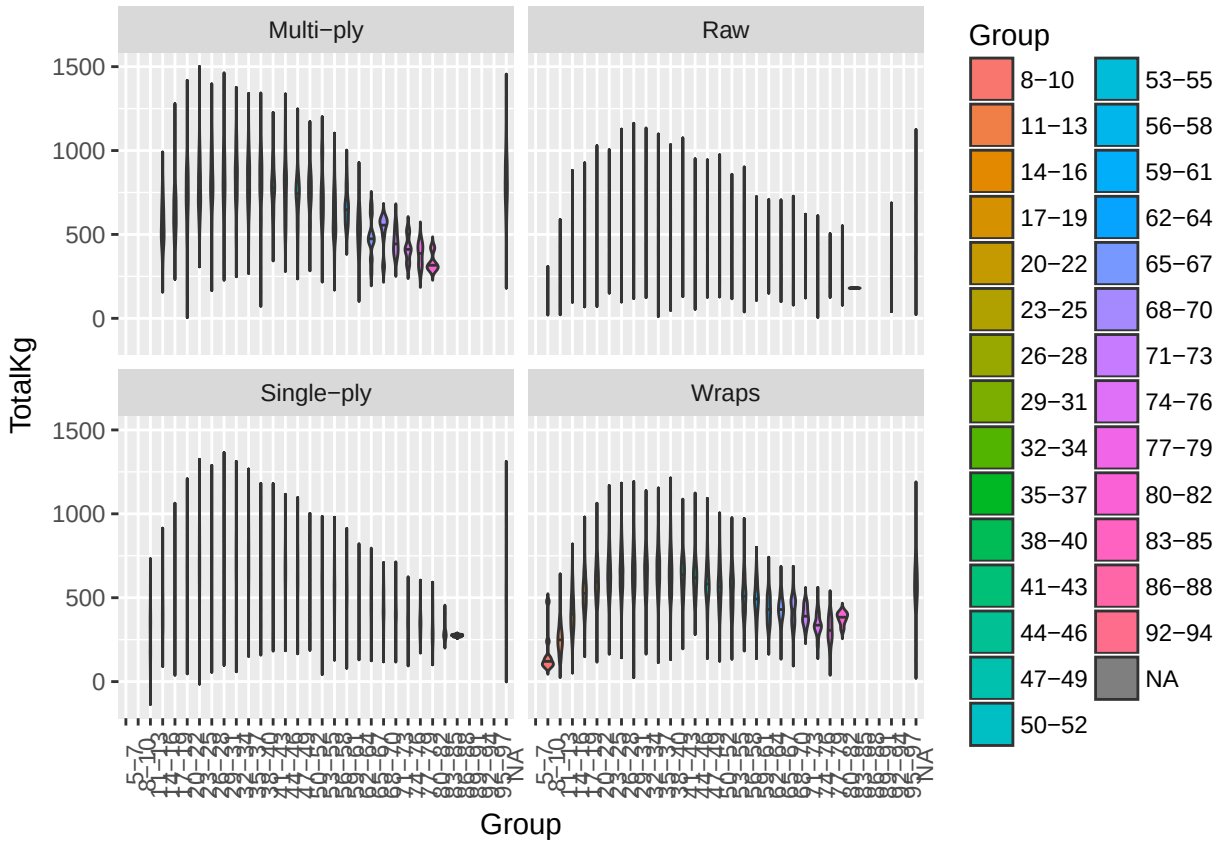

```
ggplot(women,
       aes(x=Age,y=TotalKg,color=BodyweightKg)) +
  geom_point(cex=.7) +
  facet_wrap(~Equipment) +
  scale_color_gradient2(guide="colourbar",low="yellow",mid="gold",high="red")
```

```
## Warning: Removed 44863 rows containing missing values (geom_point).
```

- There's an obvious peak around age 25, with a slowdown afterwards. What does that mean? We have a ton of data and variability, let's take a look.
- Let's see distribution of data in different ages to see how representative we are and what the resolution is

```
ggplot(men,
       aes(x=Group,y=TotalKg,fill=Group))+
  geom_violin(draw_quantiles=c(0.5),aes(fill=Group),trim=FALSE) +
  facet_wrap(~Equipment) +
  theme(axis.text.x=element_text(angle=90))
```

```
## Warning: Removed 1054 rows containing non-finite values (stat_ydensity).
```




```
# fit<-aov(TotalKg~Group,data=men)
line<-lm(TotalKg~Age+BodyweightKg+Equipment,data=men)

# layout(matrix(c(1,2,3,4),2,2))
# plot(line)
summary(line)
```

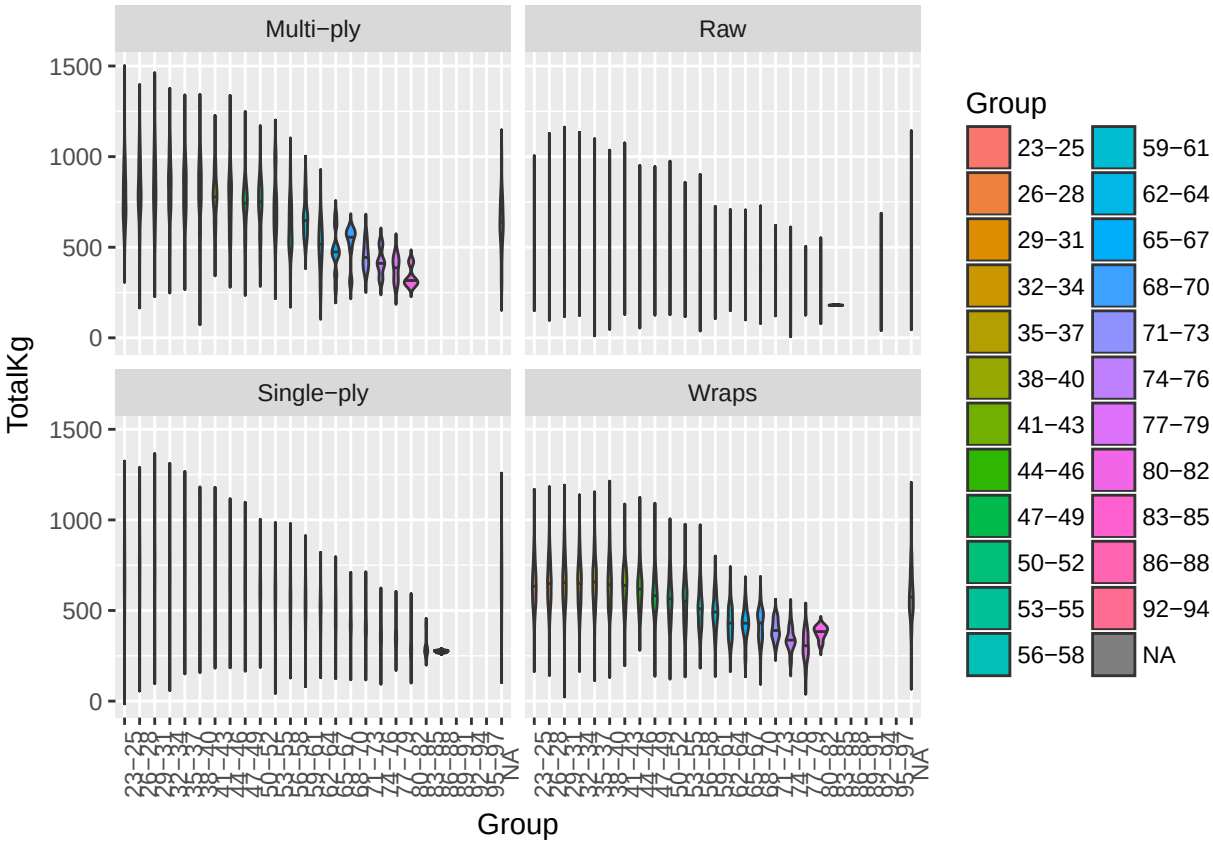
```
##
## Call:
## lm(formula = TotalKg ~ Age + BodyweightKg + Equipment, data = men)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -749.72  -67.13    4.19   72.72  509.86
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          406.64641    3.08569  131.78   <2e-16 ***
## Age                   -1.48960    0.02597  -57.36   <2e-16 ***
## BodyweightKg           3.93414    0.01526  257.86   <2e-16 ***
## EquipmentRaw        -172.96364    2.60852  -66.31   <2e-16 ***
## EquipmentSingle-ply  -63.05771    2.63002  -23.98   <2e-16 ***
## EquipmentWraps      -135.32650    2.67655  -50.56   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 111.4 on 121204 degrees of freedom
```

```
##   (176494 observations deleted due to missingness)
## Multiple R-squared:  0.4282, Adjusted R-squared:  0.4282
## F-statistic: 1.815e+04 on 5 and 121204 DF,  p-value: < 2.2e-16
```

- Let's cull the data to remove people under 22... creates a weird hump that may be obscuring some things
- Below I only work with data of men above 22 and then I'll check if this is consistent for younger males

```
men2<-men %>%
  filter(Age>22)

ggplot(men2,
       aes(x=Group,y=TotalKg,fill=Group))+
  geom_violin(draw_quantiles=c(0.5),aes(fill=Group),trim=FALSE) +
  facet_wrap(~Equipment) +
  theme(axis.text.x=element_text(angle=90))
```

```
## Warning: Removed 556 rows containing non-finite values (stat_ydensity).
```




```
line<-lm(TotalKg~Age*BodyweightKg+Equipment,data=men2)
# layout(matrix(c(1,2,3,4),2,2))
# plot(wa)
summary(line)
```

```
##
## Call:
## lm(formula = TotalKg ~ Age * BodyweightKg + Equipment, data = men2)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -757.41  -63.63    1.70   66.78  471.21
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)          5.240e+02  5.718e+00  91.643  < 2e-16 ***
## Age                 -3.136e+00  1.372e-01 -22.858  < 2e-16 ***
## BodyweightKg         3.867e+00  5.477e-02  70.599  < 2e-16 ***
## EquipmentRaw        -1.821e+02  2.588e+00 -70.364  < 2e-16 ***
## EquipmentSingle-ply -5.502e+01  2.613e+00 -21.053  < 2e-16 ***
## EquipmentWraps      -1.430e+02  2.663e+00 -53.689  < 2e-16 ***
## Age:BodyweightKg    -9.507e-03  1.484e-03  -6.406 1.51e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 103.4 on 84993 degrees of freedom
##   (654 observations deleted due to missingness)
## Multiple R-squared:  0.4896, Adjusted R-squared:  0.4896
## F-statistic: 1.359e+04 on 6 and 84993 DF,  p-value: < 2.2e-16
```

Age and Bodyweight seem to have different influences on lifting

- What's happening here
- What you see here is that when we look at Age, it has a negative relationship with Total weight lifted. What does this mean?
- One interpretation is that as you get older, it's more difficult to lift heavy weights. . . duh
- Obviously we have a bodyweight disparity here too, where the bigger you are, the more you will lift
- An interesting interaction here suggests though, that there is a negative correlation between age and bodyweight, meaning: the older you are, you lose bang for the buck in bodyweight - is this real?
- So far we have been looking at this accross all equipment types, let's break it down and see if these relationships hold among all

```
menm<-men2 %>%
  filter(Equipment=="Multi-ply")

menr<-men2 %>%
  filter(Equipment=="Raw")

mens<-men2 %>%
  filter(Equipment=="Single-ply")

menw<-men2 %>%
  filter(Equipment=="Wraps")

mm<-lm(TotalKg~Age*BodyweightKg,data=menm)
summary(mm)
```

```
##
## Call:
## lm(formula = TotalKg ~ Age * BodyweightKg, data = menm)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -759.86  -93.07    7.05   97.67  453.95
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      537.02371   59.51795   9.023  < 2e-16 ***
## Age               -4.59001    1.46269  -3.138  0.00173 **
## BodyweightKg       4.04065    0.59027   6.845 1.07e-11 ***
## Age:BodyweightKg  -0.00327    0.01481  -0.221  0.82527
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 145 on 1670 degrees of freedom
##   (18 observations deleted due to missingness)
## Multiple R-squared:  0.3416, Adjusted R-squared:  0.3405
## F-statistic: 288.9 on 3 and 1670 DF,  p-value: < 2.2e-16
```

```
mr<-lm(TotalKg~Age*BodyweightKg,data=menr)
summary(mr)
```

```
##
## Call:
## lm(formula = TotalKg ~ Age * BodyweightKg, data = menr)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -626.96  -59.04   -0.95   59.85  378.62
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      330.776514   6.962628  47.507  < 2e-16 ***
## Age               -1.716032   0.195311  -8.786  < 2e-16 ***
## BodyweightKg       3.718302   0.075538  49.224  < 2e-16 ***
## Age:BodyweightKg  -0.016672   0.002114  -7.886 3.21e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 92.56 on 40910 degrees of freedom
##   (260 observations deleted due to missingness)
## Multiple R-squared:  0.376,  Adjusted R-squared:  0.3759
## F-statistic:  8215 on 3 and 40910 DF,  p-value: < 2.2e-16
```

```
ms<-lm(TotalKg~Age*BodyweightKg,data=mens)
summary(ms)
```

```
##
## Call:
## lm(formula = TotalKg ~ Age * BodyweightKg, data = mens)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -773.28  -65.95    8.52   74.71  361.54
##
## Coefficients:
```

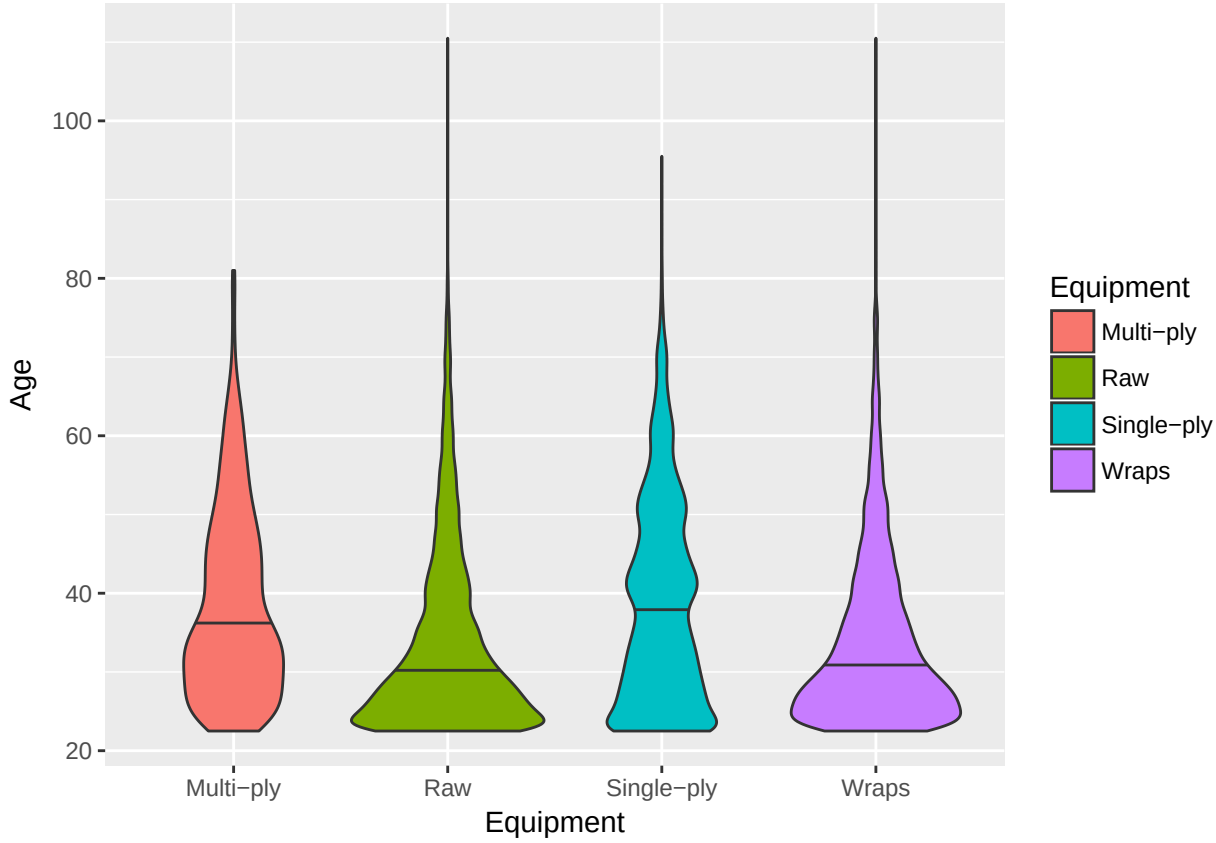
```
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)      422.415918   9.146593  46.183   <2e-16 ***
## Age               -2.935008   0.232070 -12.647   <2e-16 ***
## BodyweightKg       4.767958   0.099741  47.803   <2e-16 ***
## Age:BodyweightKg  -0.021983   0.002569  -8.556   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 113.4 on 26753 degrees of freedom
##   (283 observations deleted due to missingness)
## Multiple R-squared:  0.498,  Adjusted R-squared:  0.4979
## F-statistic:  8845 on 3 and 26753 DF,  p-value: < 2.2e-16
```

```
mw<-lm(TotalKg~Age*BodyweightKg,data=menw)
summary(mw)
```

```
##
## Call:
## lm(formula = TotalKg ~ Age * BodyweightKg, data = menw)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -748.56  -67.46   -2.63   69.28  375.32
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)      401.481627  13.173126  30.477   <2e-16 ***
## Age               -3.479179   0.372362  -9.344   <2e-16 ***
## BodyweightKg       3.538131   0.134418  26.322   <2e-16 ***
## Age:BodyweightKg  -0.002411   0.003821  -0.631    0.528
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 103 on 15651 degrees of freedom
##   (93 observations deleted due to missingness)
## Multiple R-squared:  0.3905, Adjusted R-squared:  0.3904
## F-statistic:  3343 on 3 and 15651 DF,  p-value: < 2.2e-16
```

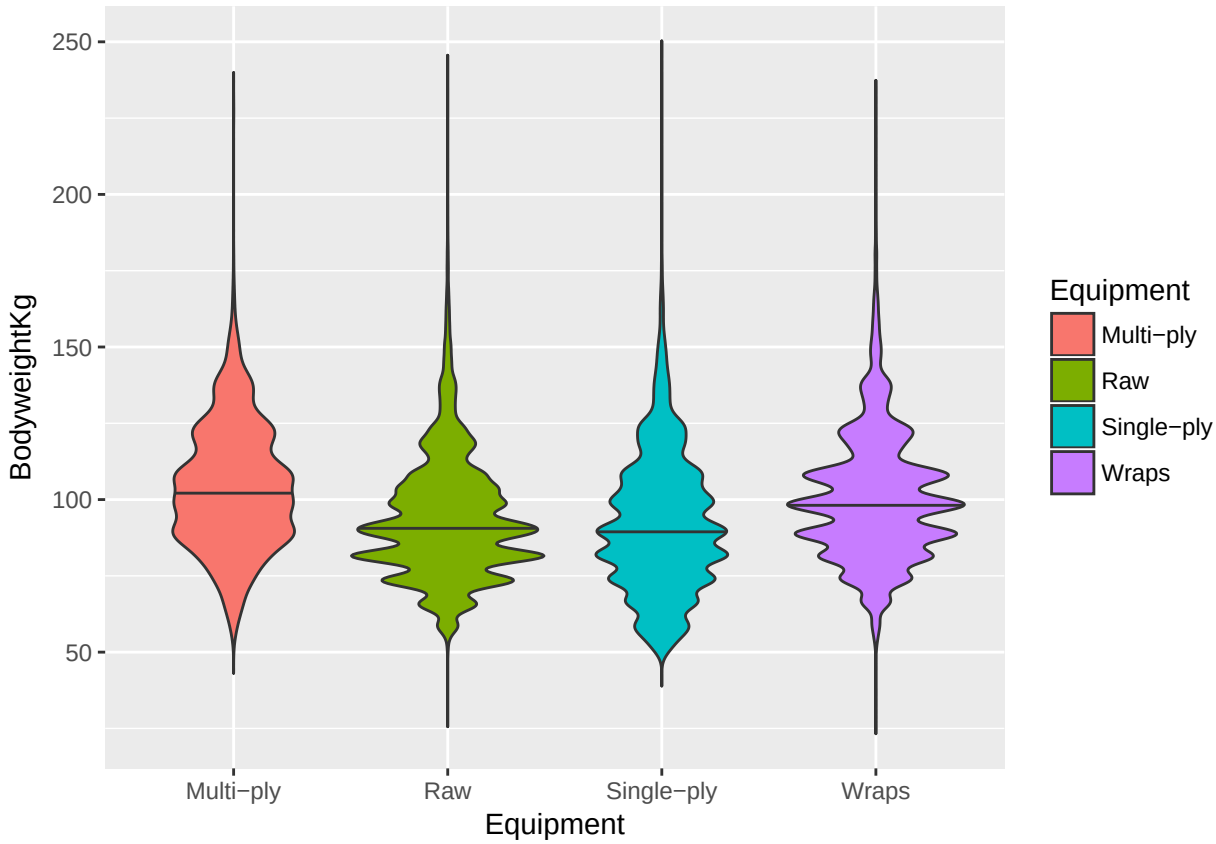
- So this is **REALLY** interesting!
- What you see here is that there is a small and significant relationship between age and total weight lifted, **but** the stronger relationship is between bodyweight and total weight lifted
- What does this mean for our interaction and the relationship between Age and Bodyweight?
- When looking at Multi-Ply lifters we see that age has a **strong negative** effect on total weight lifted, bodyweight has a **strong positive** effect... no interaction though. Meaning you don't lose weight to lifted weight relationship with age.
- Raw lifters: **Small** negative relationship of Age, positive relationship of weight, **small** but present negative interaction of age and weight (lose bang for buck of gaining weight with age)
- Single-ply: **sizeable** negative relationship of Age, **strong** positive relationship of weight, **small** but present negative interaction of age and weight
- Wraps: **Large** negative relationship of Age, positive relationship of weight, no interaction.
- is this caused just because of lack of representation of older or bigger lifters in some leagues?

```
ggplot(men2,
       aes(x=Equipment,y=Age,fill=Equipment))+
  geom_violin(draw_quantiles = c(0.5))
```




```
ggplot(men2,
       aes(x=Equipment,y=BodyweightKg,fill=Equipment))+
  geom_violin(draw_quantiles = c(0.5),trim=FALSE)
```

```
## Warning: Removed 98 rows containing non-finite values (stat_ydensity).
```



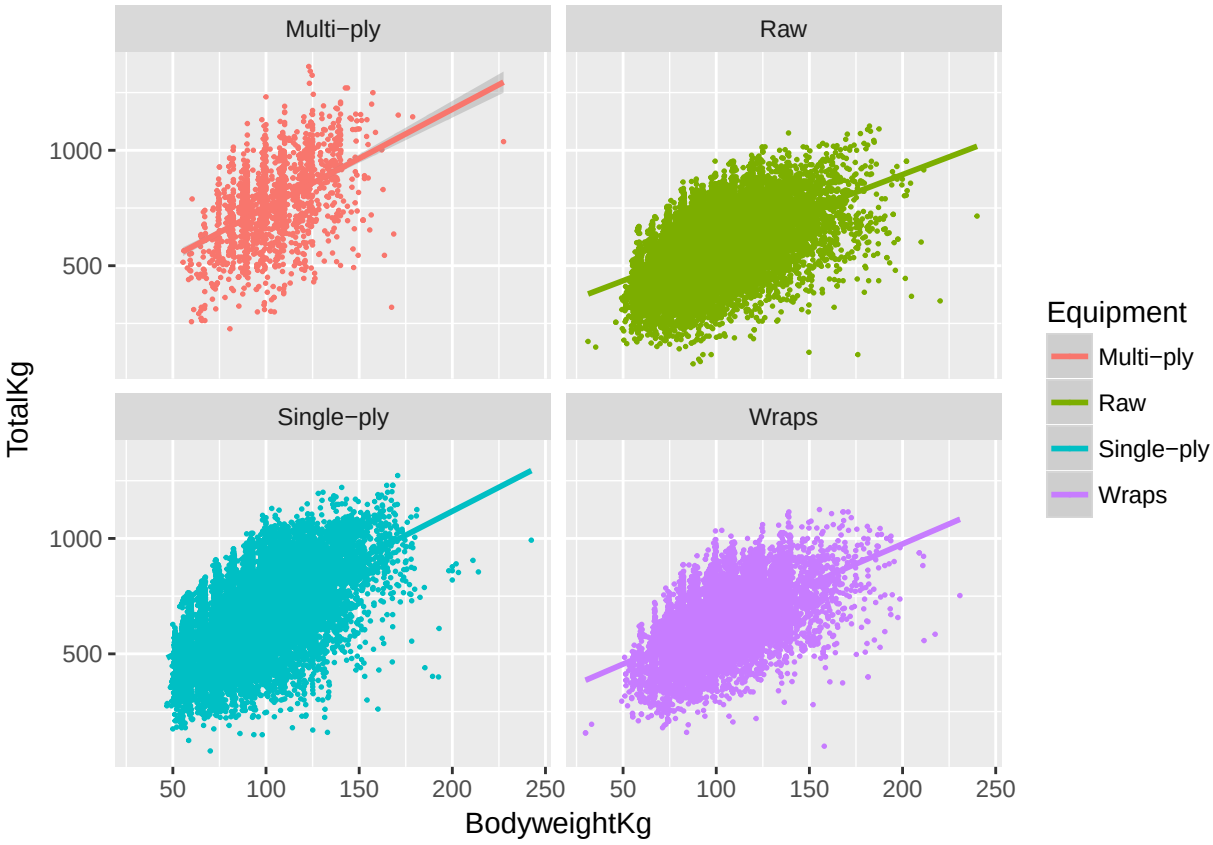

```
ggplot(men2,
       aes(x=Age,y=BodyweightKg,color=Equipment))+
  geom_point()+
  facet_wrap(~Equipment)
```

```
## Warning: Removed 98 rows containing missing values (geom_point).
```

- When we look at Age, we do see that there are proportionally slightly more older lifters in the Multy-ply and Single-ply categories
- There is a strange trend to have lifters of heavier lifters in the Multi-ply and Wraps divisions. This could be a difference in the weight categories offered by different leagues that support multi-ply, vs ones that don't.
- you can see clearly the categories by the wiggle in distribution, each wiggle is a high proportion of people likely at the top of their weight class, where there are less lifters at the bottom of a weight class +These are the only two divisions that do not have a interaction between age and weight
- Overall I am not super convinced by the bodyweight and age interaction term. It seems like there is some skews to the data that could be putting less heavy lifters in the older categories, while there is high overrepresentation of young lifters there (Raw, Wraps).
- On the other hand, the large effect size of Age and Bodyweight separately and in opposite directions is consistent among all categories. It is strange that the effect of age is **much smaller** in raw and single-ply categories.
- Something smells fishy here though! (and no it's not the knee sleeves I haven't washed in 3 months)
- We saw that both Age and Weight have some effect on the weight lifted by people, but do we just see a large proportion of monster heavyweight lifters in the older categories?
- Let me clarify here - big guys lift big weights (generally) -> case and point

```
ggplot(men2,
       aes(x=BodyweightKg,y=TotalKg,color=Equipment))+
  geom_point(cex=.3)+
  facet_wrap(~Equipment) +
  geom_smooth(method=lm)
```

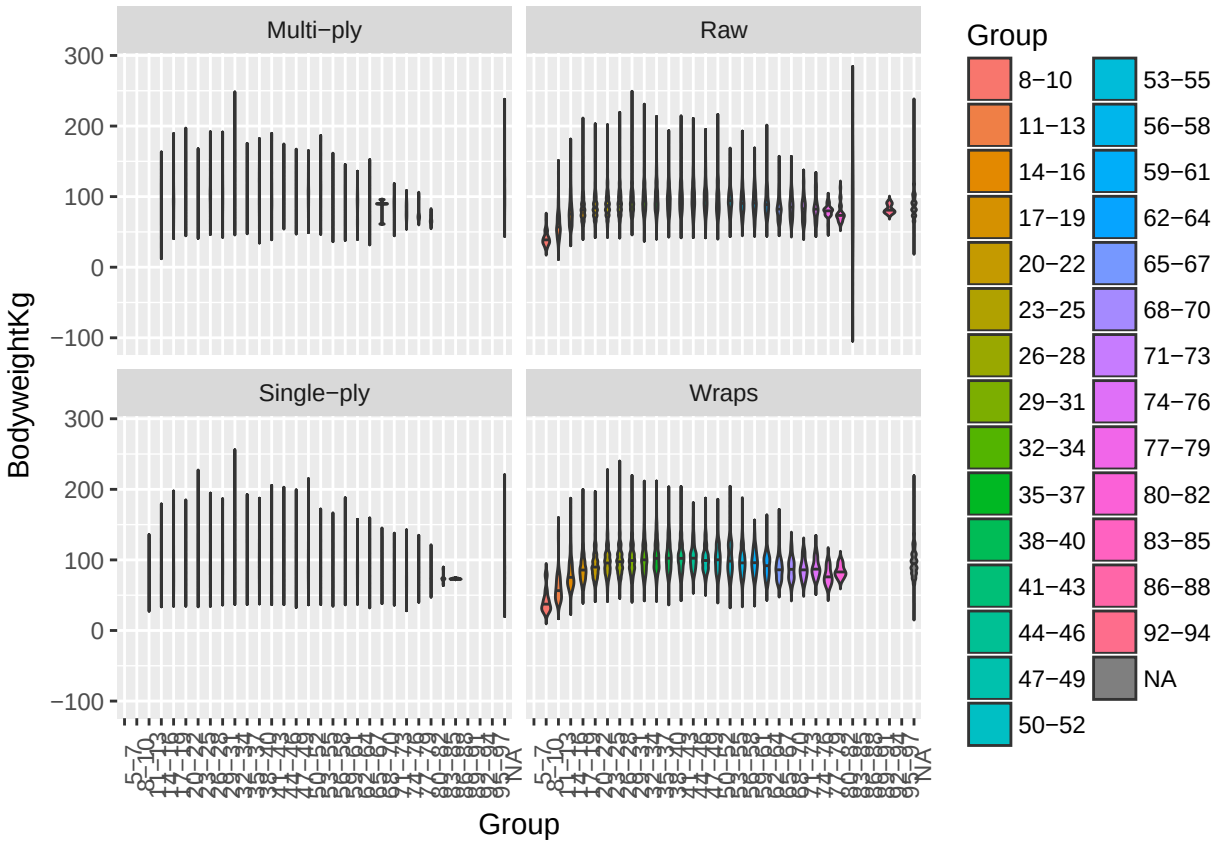
```
## Warning: Removed 654 rows containing non-finite values (stat_smooth).
## Warning: Removed 654 rows containing missing values (geom_point).
```



- There's fewer big guys in the higher age ranges

```
ggplot(men,
       aes(x=Group,y=BodyweightKg,color=TotalKg))+
  geom_violin(draw_quantiles=c(0.5),aes(fill=Group),trim=FALSE)+
  scale_color_gradient2(guide="colourbar",low="yellow",mid="gold",high="red") +
  facet_wrap(~Equipment) +
  theme(axis.text.x=element_text(angle=90))
```

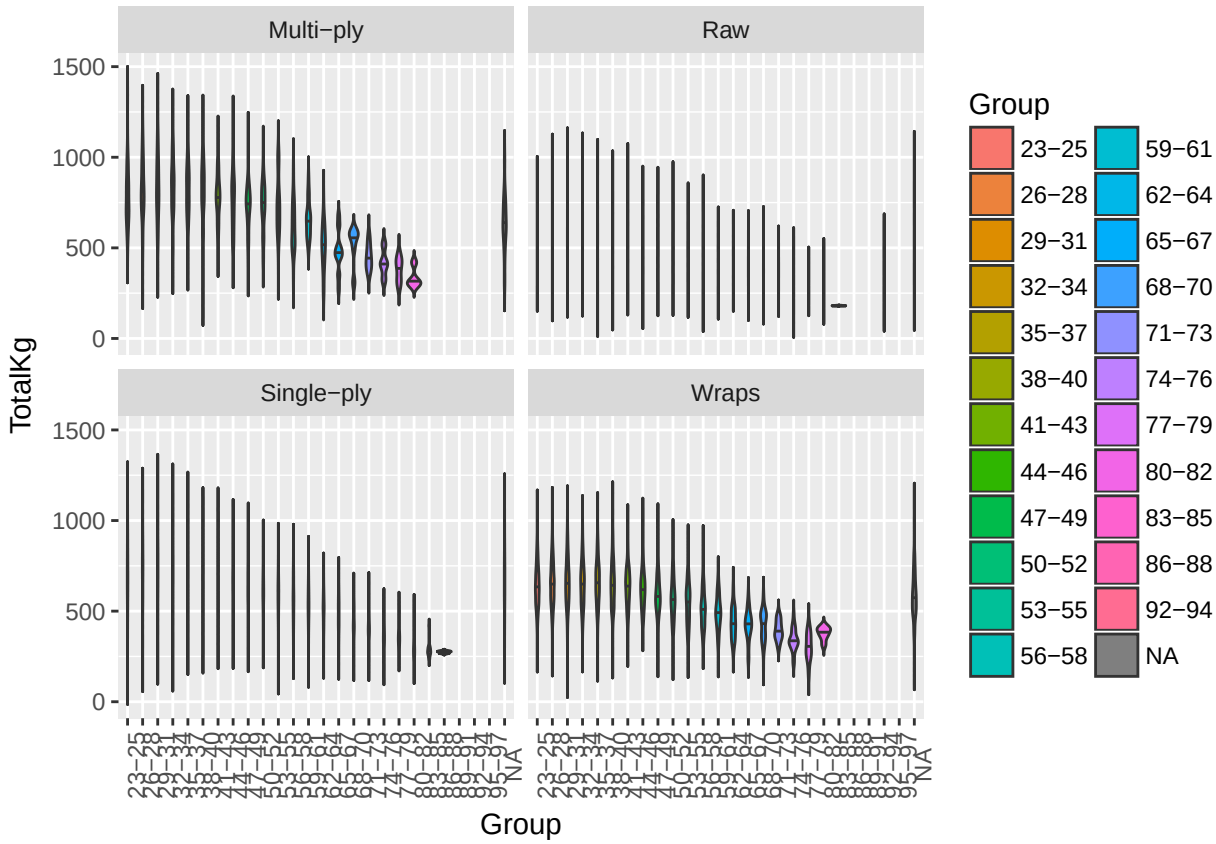
```
## Warning: Removed 1037 rows containing non-finite values (stat_ydensity).
```



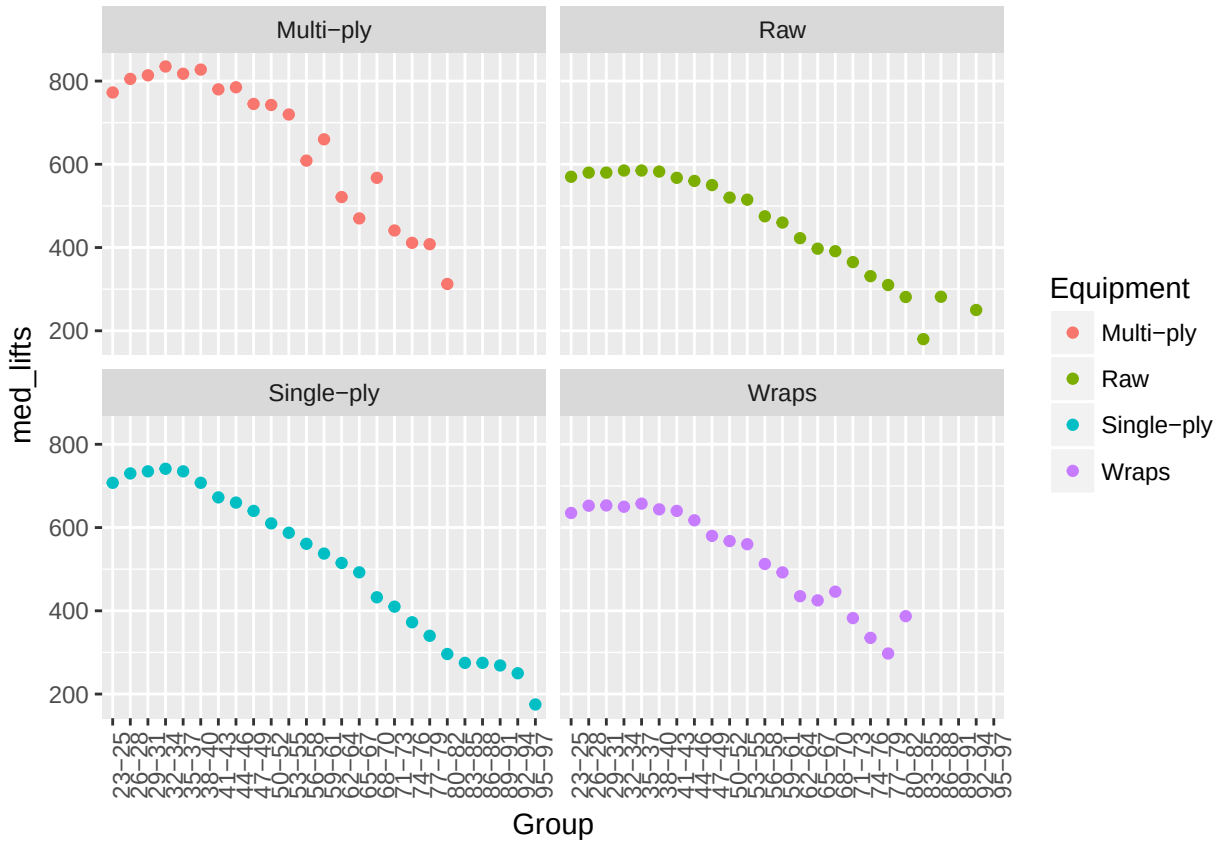
- **BUT** I think it becomes obvious that these relationships behave very differently if we talk about means and maximums. These two metrics paint two separate pictures and I think bodyweight may be part of the explanation (at least in this dataset)
- What you see below is the median (less affected by skew than the mean) total weight lifted in each weight category, as well as the maximum weight lifted

```
ggplot(men2,
       aes(x=Group,y=TotalKg,fill=Group))+
  geom_violin(draw_quantiles=c(0.5),aes(fill=Group),trim=FALSE) +
  facet_wrap(~Equipment) +
  theme(axis.text.x=element_text(angle=90))
```

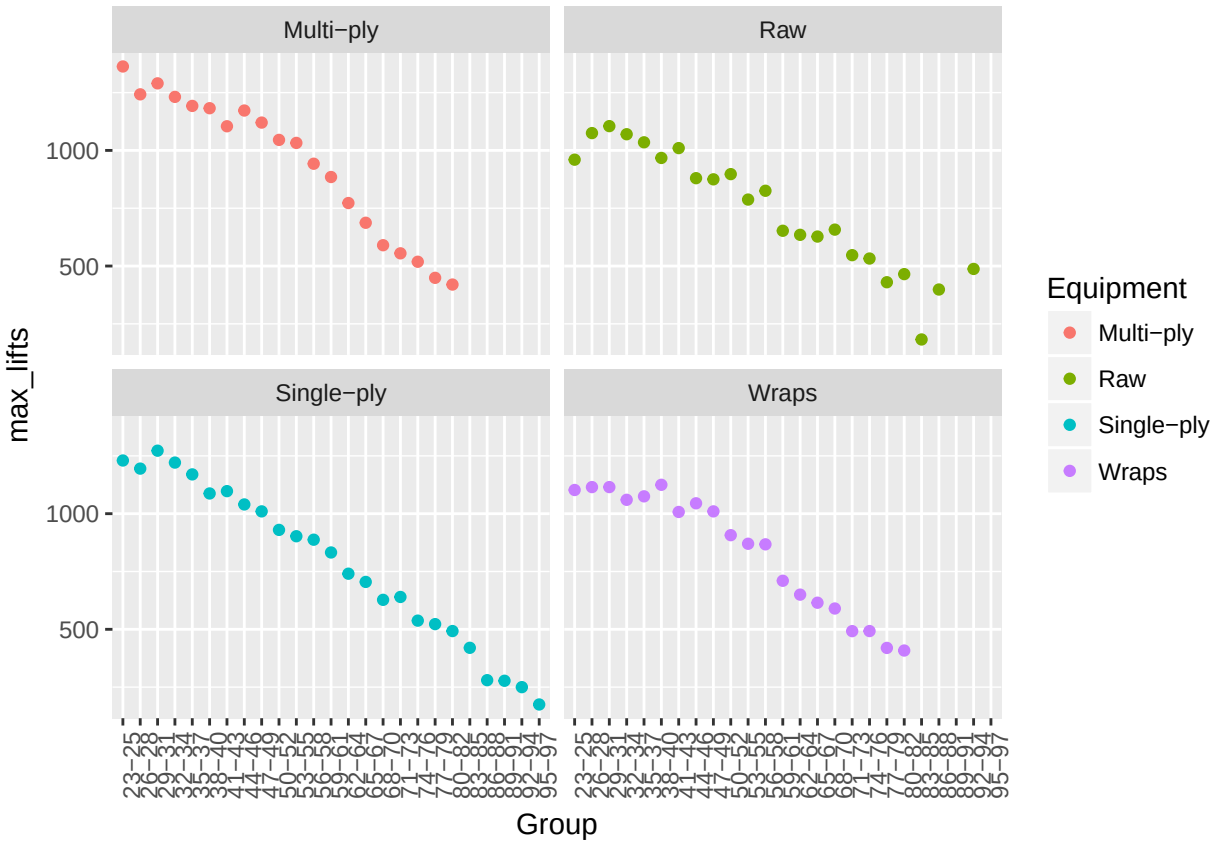
```
## Warning: Removed 556 rows containing non-finite values (stat_ydensity).
```




```
menmed<-men2 %>%
  filter(!is.na(Group),!is.na(BodyweightKg),!is.na(TotalKg)) %>%
  group_by(Group,Equipment) %>%
  summarize(med_lifts=median(TotalKg))

menmax<-men2 %>%
  filter(!is.na(Group),!is.na(BodyweightKg),!is.na(TotalKg)) %>%
  group_by(Group,Equipment) %>%
  summarize(max_lifts=max(TotalKg))

ggplot(menmed,
       aes(x=Group,y=med_lifts,color=Equipment))+
  geom_point()+
  facet_wrap(~Equipment)+
  theme(axis.text.x=element_text(angle=90))
```




```
ggplot(menmax,
       aes(x=Group,y=max_lifts,color=Equipment))+
  geom_point()+
  facet_wrap(~Equipment)+
  theme(axis.text.x=element_text(angle=90))
```



- What you see is that the maximum weight lifted in each category declines much more rapidly than the average weight. This does differ between Equipment type with Raw being the smoothest decline and Multi-ply seeming to be the strongest
- as example let's take the difference between average 23-25 category and 50-52

```
med<-menmed %>%
  filter(Group=="23-25" || Group=="50-52")

max<-menmax %>%
  filter(Group=="23-25" || Group=="50-52")

meddif<-c()
for(i in 1:(dim(med)[[1]]/2)){
  meddif[i]<-med[i,"med_lifts"]-med[i+4,"med_lifts"]
}

maxdif<-c()
for(i in 1:(dim(max)[[1]]/2)){
  maxdif[i]<-max[i,"max_lifts"]-max[i+4,"max_lifts"]
}
equip<-c("Multi-ply","Raw","Single-ply","Wraps")
names(meddif)<-equip
names(maxdif)<-equip
print(meddif)
```

```
## $`Multi-ply`
## [1] 29.99
##
```

```
## $Raw
## [1] 50
##
## $`Single-ply`
## [1] 97.5
##
## $Wraps
## [1] 67.51
```

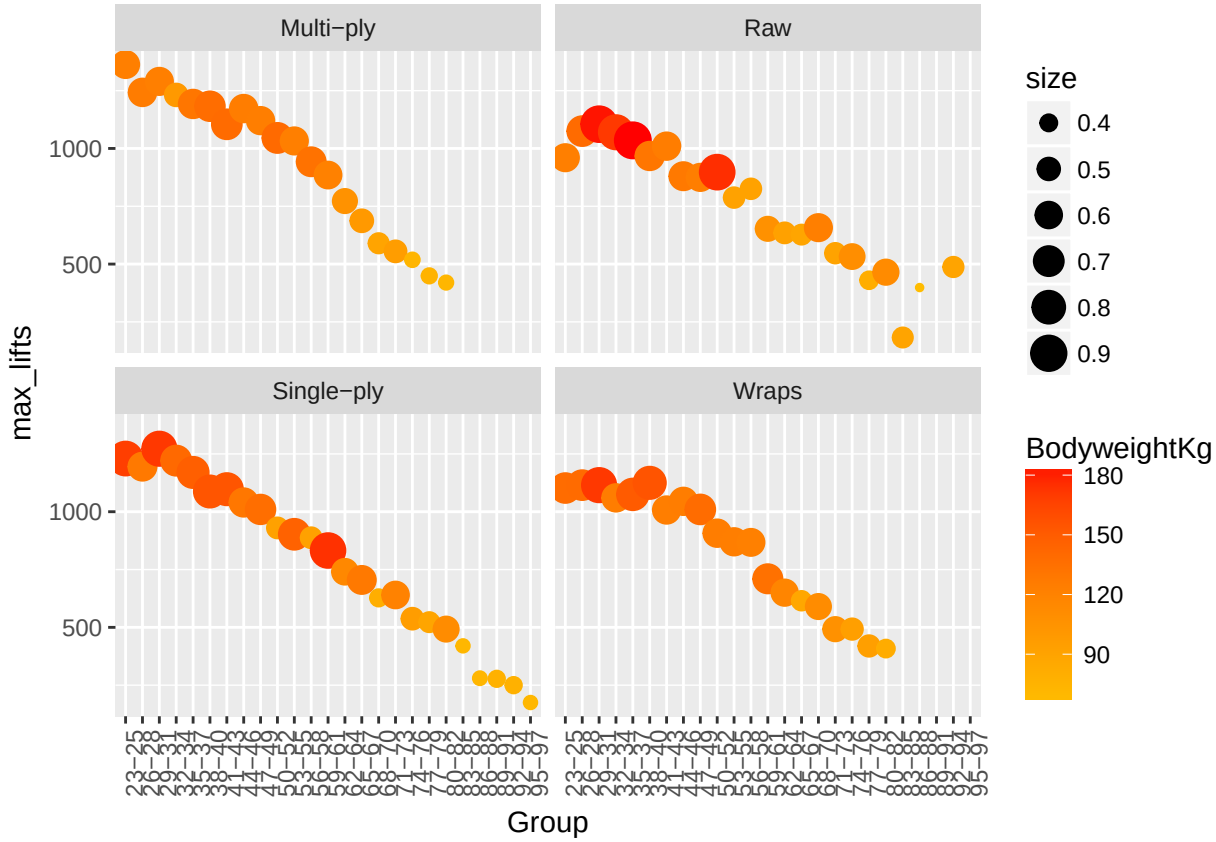
```
print(maxdif)
```

```
## $`Multi-ply`
## [1] 317.52
##
## $Raw
## [1] 62.5
##
## $`Single-ply`
## [1] 300
##
## $Wraps
## [1] 195.32
```

- What you see here are the differences between median and then maximum weight lifted at those two age categories.
- when we look at median, you see <100kg difference in all categories, and it is as small as 50 kg difference in mean total weight lifted in the raw division
- when we look at maximums - **large** differences appear!
- Raw is still with the smallest difference of 62.5kg, while Multi-ply lifters lift 317.52 kg more as a maximum of the 23-25kg category than the maximum in the 50-52. **WHY**?
- Let's resubset our data and see who is lifting those weights and how much do they weigh compared to their max counterparts in the older categories

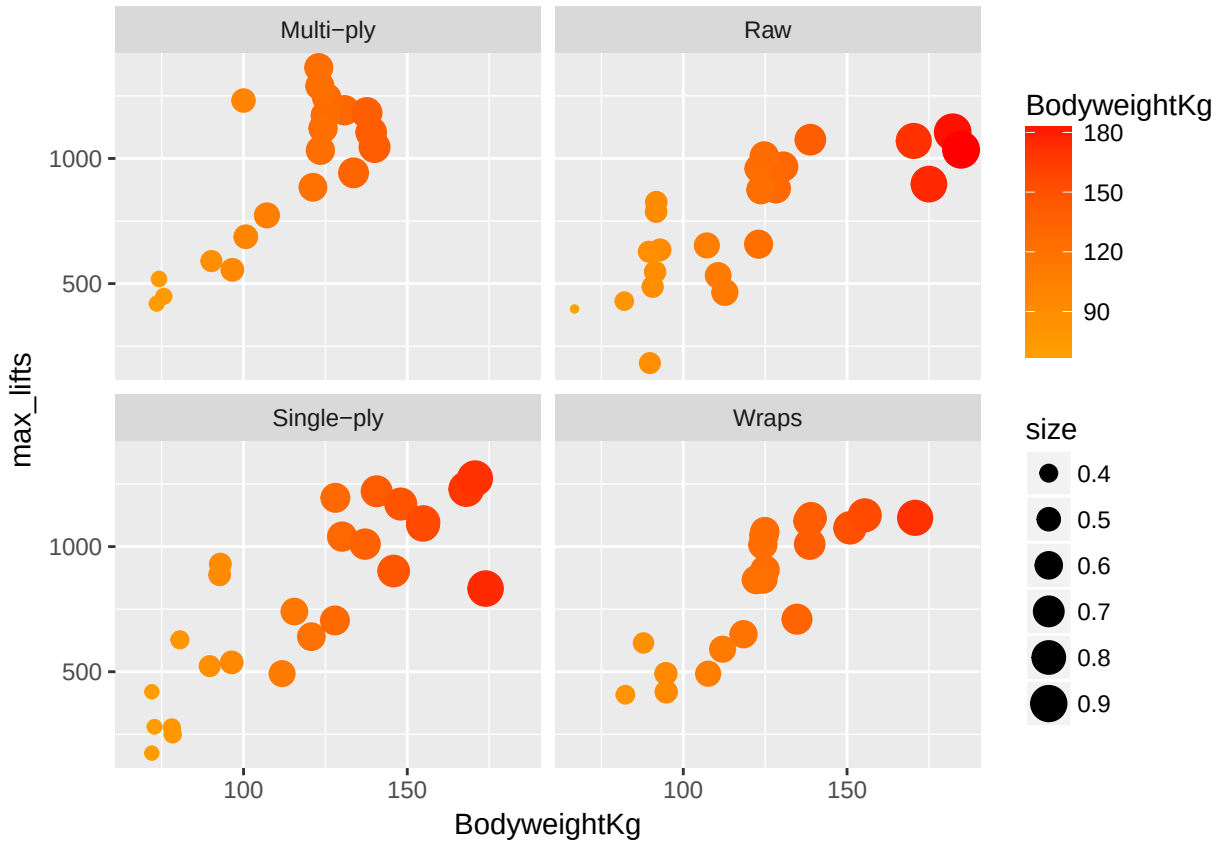
```
menmax<-men2 %>%
  filter(!is.na(Group),!is.na(BodyweightKg),!is.na(TotalKg)) %>%
  group_by(Group,Equipment,BodyweightKg) %>%
  summarize(max_lifts=max(TotalKg))

equip<-c("Multi-ply","Raw","Single-ply","Wraps")

max<-c()
for(i in unique(menmax$Group)){
  for(j in equip){
  sub<-menmax %>%
    filter(Group==i && Equipment==j)
  ord<-order(sub$max_lifts,decreasing=TRUE)
  sub2<-sub[ord,]
  max<-rbind(max,sub2[1,])
}
}

max2<-max %>%
  filter(!is.na(Group)) %>%
  mutate(size=BodyweightKg/200)

ggplot(max2,
```

```
      aes(x=Group,y=max_lifts,color=BodyweightKg))+
  geom_point(aes(cex=size))+
  facet_wrap(~Equipment)+
  theme(axis.text.x=element_text(angle=90))+
  scale_color_gradient2(guide="colourbar",low="yellow",mid="yellow",high="red")
```



- If that doesn't show it, what does. The size and redness of the dot is the weight of the lifter who made a maximum lift. As you get to the older categories, what you see is a consistent decline in both maximum lift and in weight of the lifter.
- Let's show this a little more explicitly with a relationship between weight lifted and weight of lifter

```
ggplot(max2,
       aes(x=BodyweightKg,y=max_lifts,color=BodyweightKg))+
  geom_point(aes(cex=size))+
  facet_wrap(~Equipment)+
  scale_color_gradient2(guide="colourbar",low="yellow",mid="gold",high="red")
```

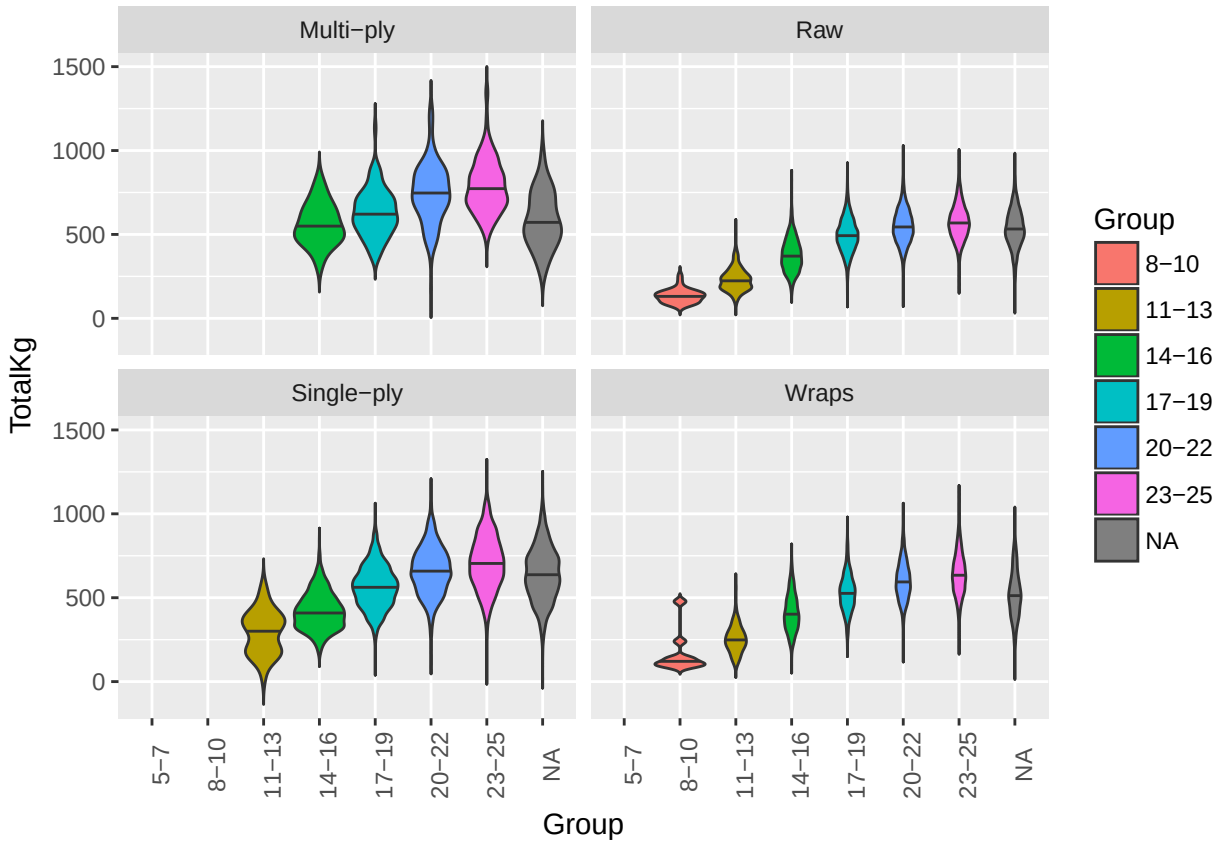


Younglings

- Let's verify why I threw out the data of people under 22.
- They create a hump in the distribution of Total weight lifted because they are still getting stronger. If we plot these side by side, you can see that lifters get stronger till about 23-25 and then plateau there.

```
men3<-men %>%
  filter(Age<26)

ggplot(men3,
       aes(x=Group,y=TotalKg,fill=Group))+
  geom_violin(draw_quantiles=c(0.5),aes(fill=Group),trim=FALSE) +
  facet_wrap(~Equipment) +
  theme(axis.text.x=element_text(angle=90))
```

```
## Warning: Removed 230 rows containing non-finite values (stat_ydensity).
```




```
ym<-lm(TotalKg~Age*BodyweightKg+Equipment,data=men3)
summary(ym)
```

```
##
## Call:
## lm(formula = TotalKg ~ Age * BodyweightKg + Equipment, data = men3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -675.43  -59.20    1.71   62.01  492.13
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -2.131e+01  1.210e+01  -1.761   0.0783 .
## Age                   1.874e+01  5.410e-01  34.640   <2e-16 ***
## BodyweightKg          3.873e+00  1.313e-01  29.497   <2e-16 ***
## EquipmentRaw         -1.415e+02  4.642e+00 -30.486   <2e-16 ***
## EquipmentSingle-ply -3.863e+01  4.675e+00  -8.262   <2e-16 ***
## EquipmentWraps       -1.153e+02  4.720e+00 -24.418   <2e-16 ***
## Age:BodyweightKg     -1.707e-02  6.259e-03  -2.727   0.0064 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 95.6 on 57093 degrees of freedom
##   (298 observations deleted due to missingness)
## Multiple R-squared:  0.5481, Adjusted R-squared:  0.548
## F-statistic: 1.154e+04 on 6 and 57093 DF,  p-value: < 2.2e-16
```

- The linear model here reveals a strong positive relationship of both Age and Weight, but a small and barely significant relationship of the interaction term. This means that gaining weight at age 15 gives you slightly more advantage in your total lifts than gaining weight at 23, but slightly!

Federation

- Let's take a look at the meets.csv file which gives us some more data on federations in which these occurred.

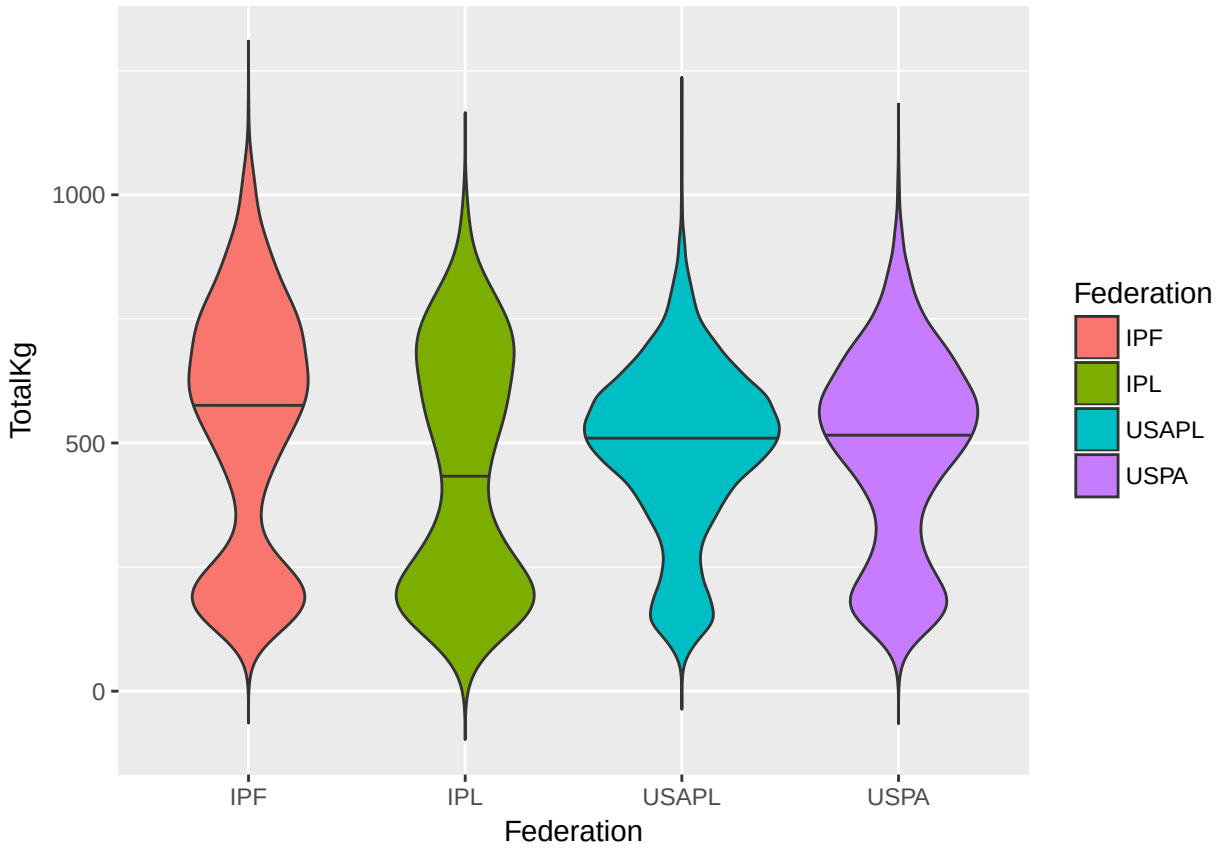
```
meets<-read.csv("~/power/data/open/openpowerlifting/meets.csv")
str(meets)
```

```
## 'data.frame':    12064 obs. of  8 variables:
##  $ MeetID     : int  0 1 2 3 4 5 6 7 8 9 ...
##  $ MeetPath   : Factor w/ 12064 levels "365strong/1601",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ Federation : Factor w/ 83 levels "365Strong","AAPF",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ Date       : Factor w/ 3672 levels "1971-01-01","1971-04-18",..: 3348 3363 3281 3266 3229 3472 340
##  $ MeetCountry: Factor w/ 83 levels "Algeria","Argentina",..: 80 80 80 80 80 80 80 80 80 80 ...
##  $ MeetState  : Factor w/ 90 levels "","AB","ACT",..: 47 42 47 71 71 47 21 42 71 47 ...
##  $ MeetTown   : Factor w/ 2149 levels "","Aalborg","Aalborg SK",..: 316 1385 316 1565 1565 316 1136 9
##  $ MeetName   : Factor w/ 6299 levels "015 Pennsylvania State Bench Press and Deadlift",..: 2569 5510
```

- Let's merge these two databases

```
all<-merge(open,meets,by="MeetID")

men<-all %>%
  filter(Sex=="M") %>%
  filter(Federation==c("USAPL","USPA","IPF","IPL"))

ggplot(men,
       aes(x=Federation,y=TotalKg,fill=Federation))+
  geom_violin(draw_quantiles = 0.5,trim=FALSE)
```

```
## Warning: Removed 2049 rows containing non-finite values (stat_ydensity).
```


1000
500
0
TotalKg
IPF
IPL
USAPL
USPA
Federation
Federation
IPF
IPL
USAPL
USPA