A Statistical Analysis of Power Lifting in Canada, the USA, and the world

Ousman-Jikineh

2024-02-18

```
library(tidyverse)
```

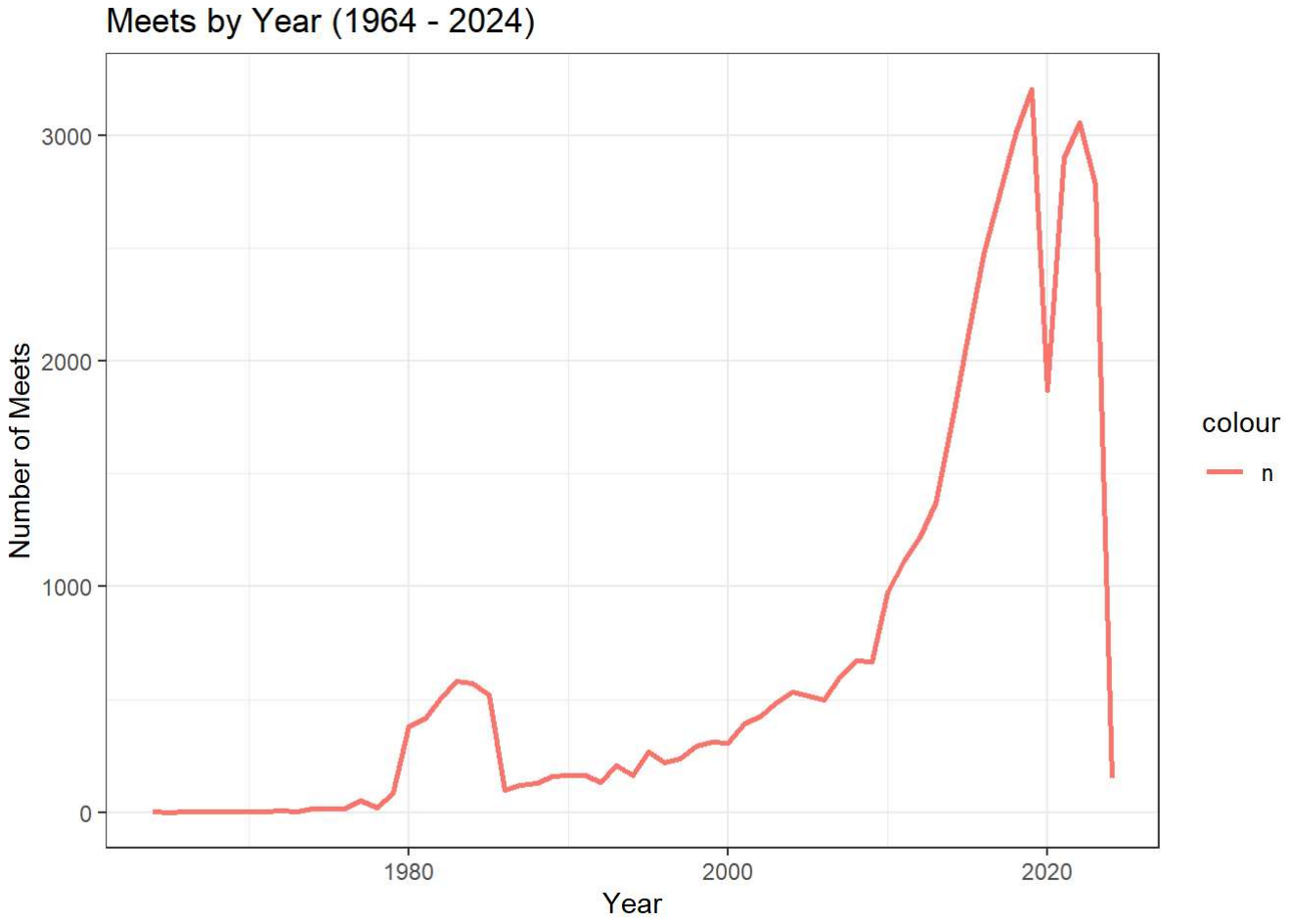
```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.4     ✓ readr     2.1.5
## ✓ forcats   1.0.0     ✓ stringr   1.5.1
## ✓ ggplot2   3.4.4     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.1
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to
become errors
```

```
data <- read.csv("C:/Users/ser/Downloads/Projects/R/openpowerlifting_large/pl_data.csv")
data = data %>% mutate(Year = as.integer(str_sub(Date,1,4)))
Year_summary <- data %>% group_by(Year) %>% summarise(Meet_Count= n_distinct(MeetName, na.
rm=TRUE))

g <- ggplot(Year_summary, aes(x=Year, y=Meet_Count)) + geom_line(size=1, aes(col="n")) +
  labs(title = "Meets by Year (1964 - 2024)", x = "Year", y="Number of Meets") + theme_bw
()
```

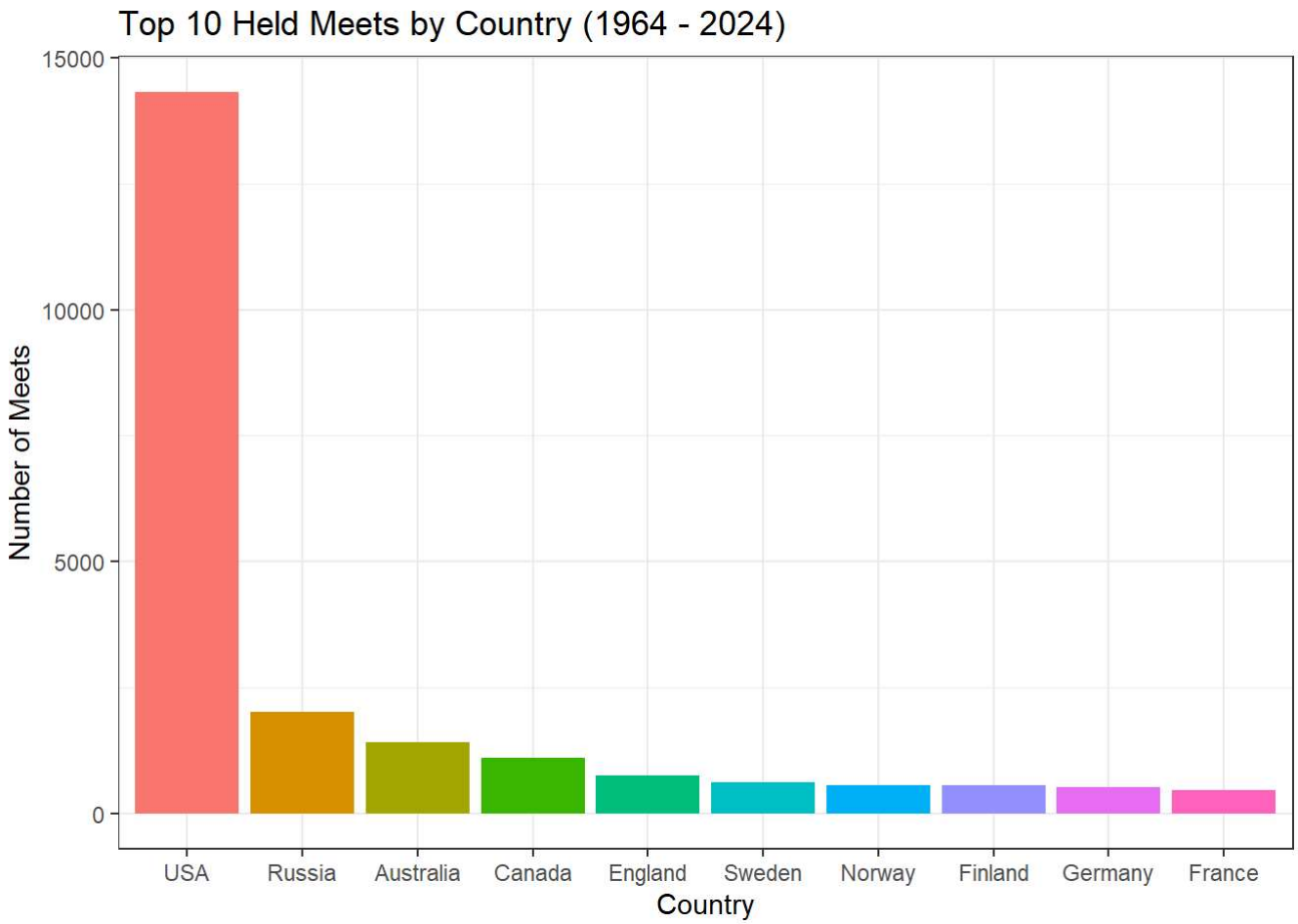
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
g
```



This data set contains information dating from the 60s. There was a noticeable jump in popularity in the early 80's, likely due to the fact the power lifting only became recognized as a sport in the 70s. However, power lifting didn't become very popularized until the 2010's.

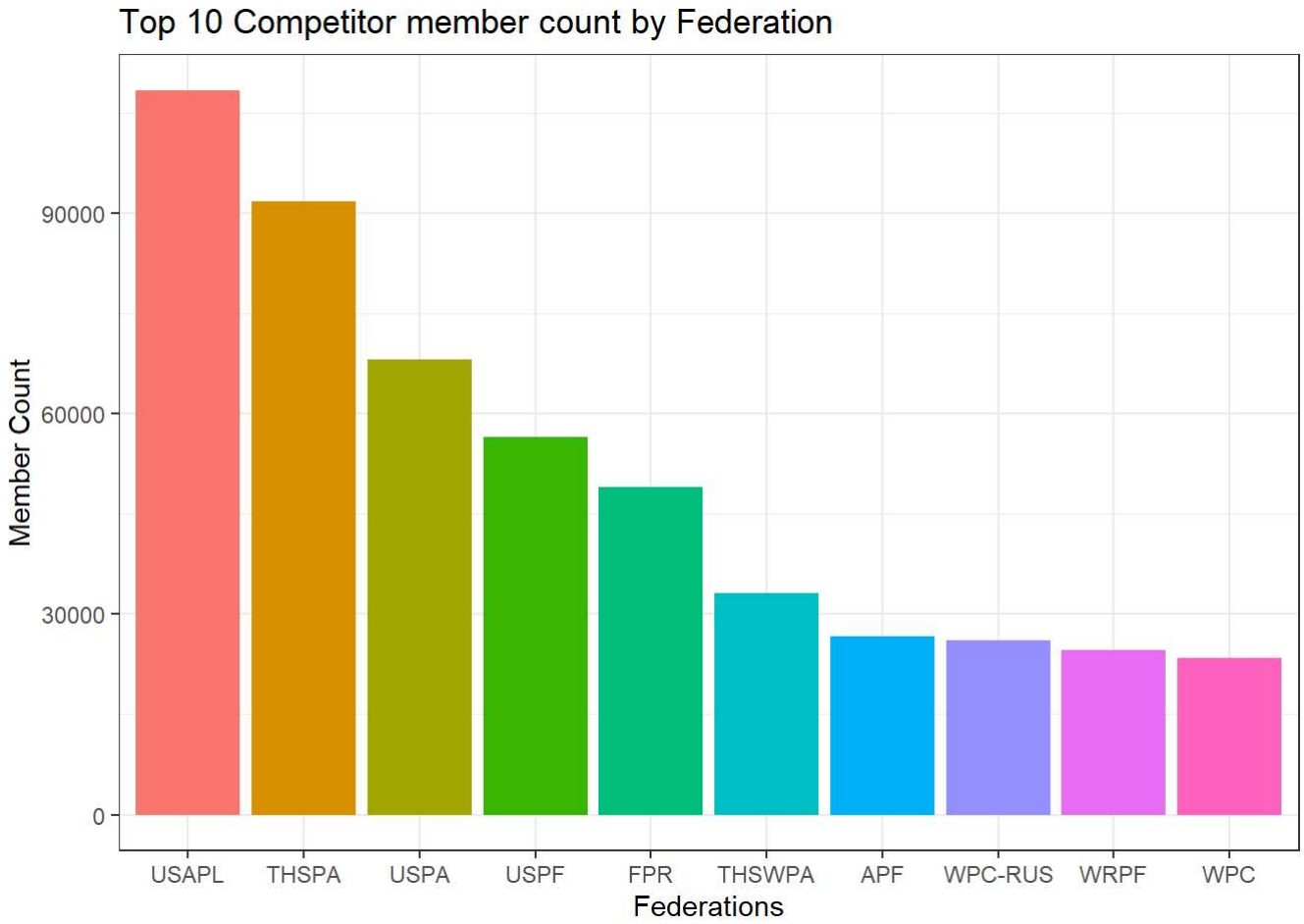
```
unique_meets <- data %>%
  distinct(Year, MeetCountry, MeetName)
summary_data <- unique_meets %>%
  group_by(MeetCountry) %>%
  summarise(TotalMeets = n_distinct(MeetName)) %>% arrange(desc(TotalMeets)) %>%
  mutate(MeetCountry = factor(MeetCountry, levels = MeetCountry)) %>% slice(1:10)
g <- ggplot(summary_data, aes(x=MeetCountry, y = TotalMeets, fill=MeetCountry)) + geom_bar
(stat="identity") +
  labs(title="Top 10 Held Meets by Country (1964 - 2024)", x = "Country", y="Number of Mee
ts") + theme_bw() + theme(legend.position = "none")
g
```



There is large bias towards English speaking countries, with the exception beign Russia coming in second.

```
pop_federation <- data %>% group_by(Federation) %>% summarise(members = n_distinct(Name))
top_10_federations <- pop_federation %>%
  arrange(desc(members)) %>%
  mutate(Federation = factor(Federation, levels = Federation)) %>%
  slice(1:10)

plot_federation <- ggplot(top_10_federations, aes(x=Federation, y=members, fill=Federatio
n)) +
  geom_bar(stat = "identity") +
  labs(title = "Top 10 Competitor member count by Federation", x="Federations", y = "Membe
r Count") + theme_bw() + theme(legend.position = "none")
plot_federation
```



Once again there is a clear bias towards the USA and English speaking countries. This makes sense considering that power lifting originated in the United States and the UK.

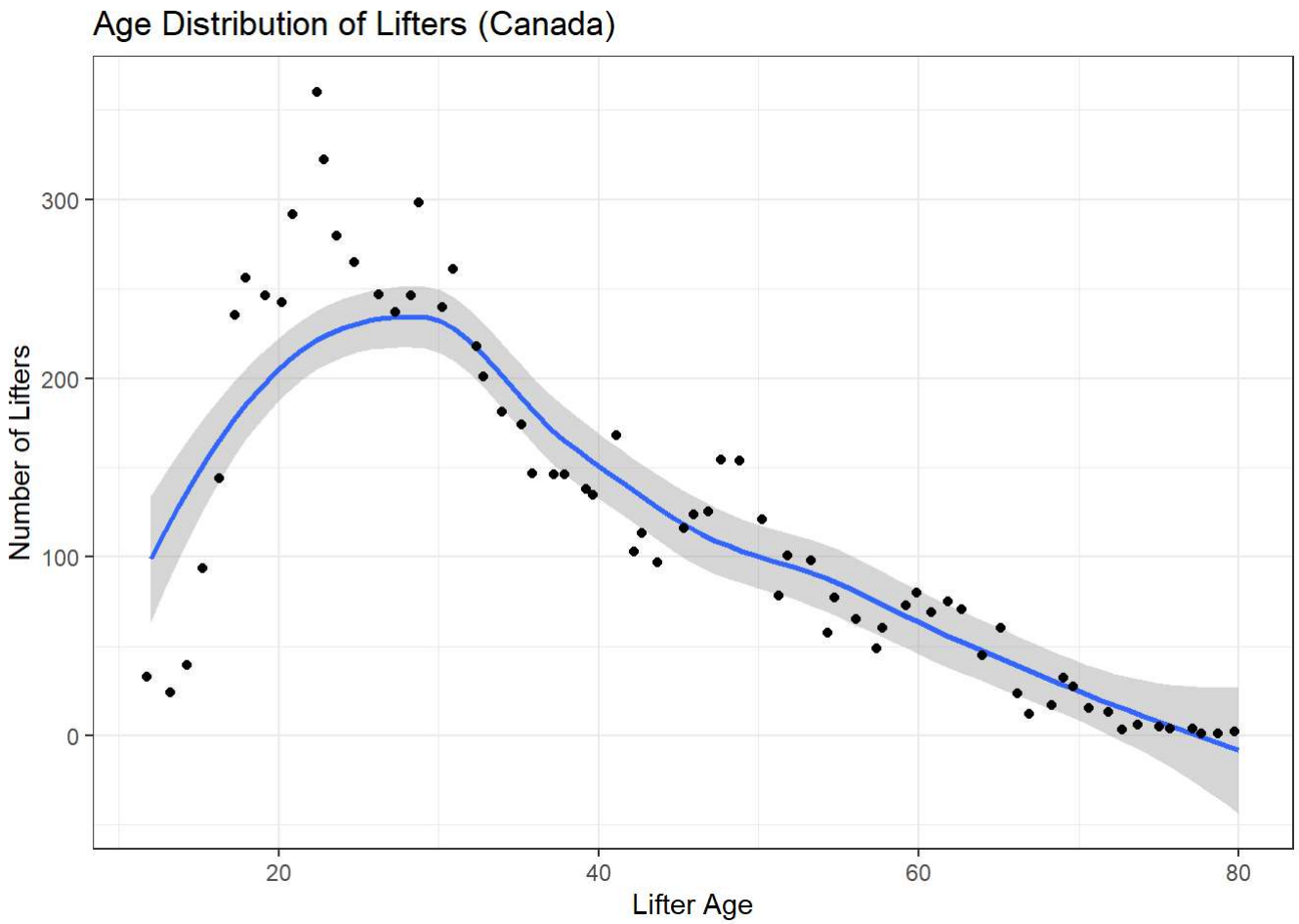
Power Lifting Meets in Canada and the USA, from 2020 - 2024

From this point, I will be filtering the data for competitors ages 12 and up. This is in order to make the data more relevant for my comparison. ### What is the age distribution of Lifters in the US and CANADA

```
library(tidyverse)
data = data %>% drop_na(Age) %>% mutate(Age = as.integer(Age)) %>%
  filter(as.integer(str_extract(Date, "\\d{4}")) >= 2020, Age >= 12)
canada_meets = data %>% filter(MeetCountry == "Canada")

age_dist_CAN <- canada_meets %>% drop_na(Age) %>% group_by(age = as.integer(Age)) %>% summ
arise(count = n())
g <- ggplot(age_dist_CAN, aes( x = age, y=count)) + geom_smooth() + geom_jitter() +
  labs(title = "Age Distribution of Lifters (Canada)", x="Lifter Age", y="Number of Lifter
s") + theme_bw()
g
```

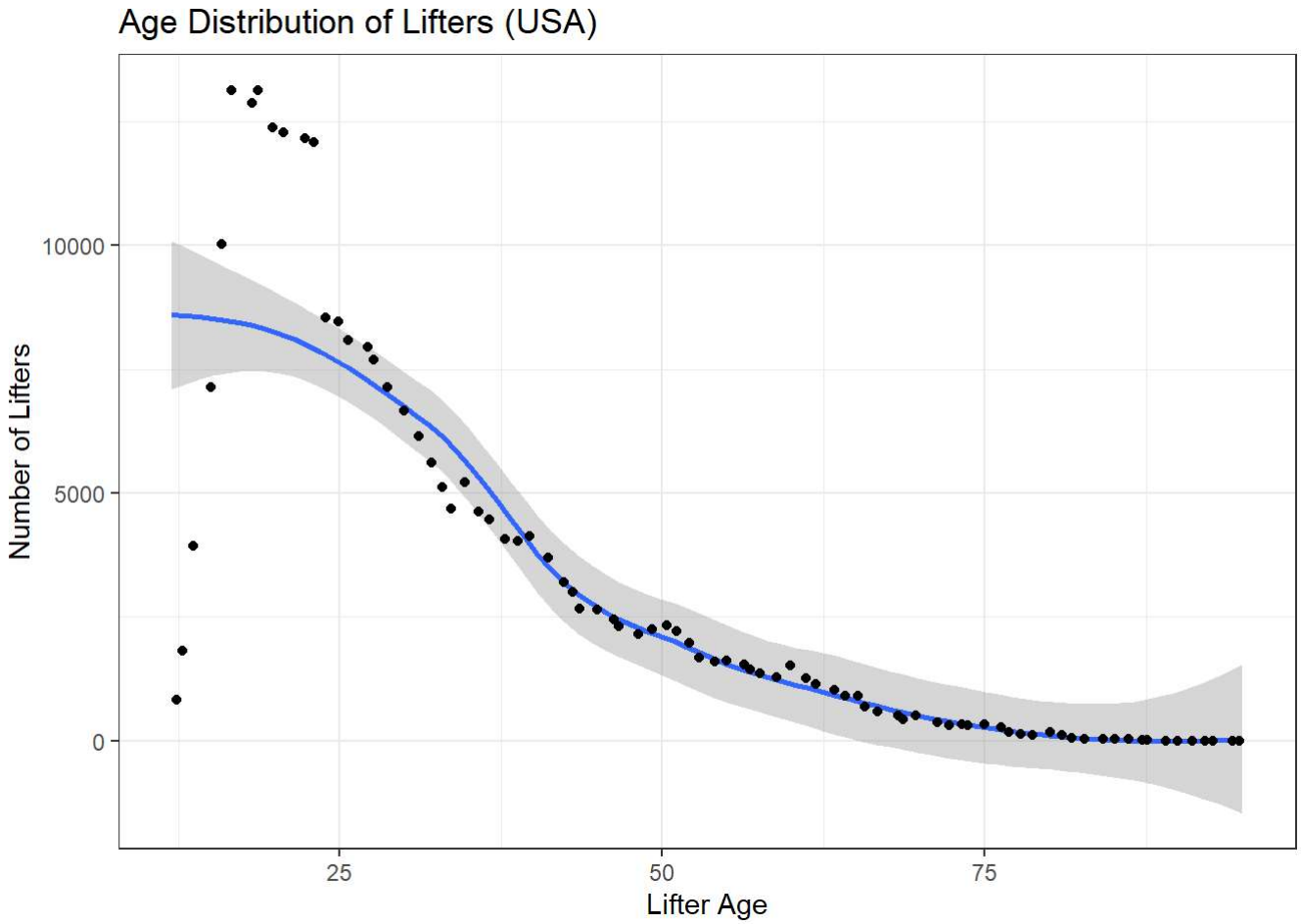
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
usa_meets = data %>% filter(MeetCountry == "USA")

age_dist_USA <- usa_meets %>% drop_na(Age) %>% group_by(age = as.integer(Age)) %>% summari
se(count = n())
g <- ggplot(age_dist_USA, aes( x = age, y=count)) + geom_smooth() + geom_jitter() +
  labs(title = "Age Distribution of Lifters (USA)", x="Lifter Age", y="Number of Lifters")
+ theme_bw()
g
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



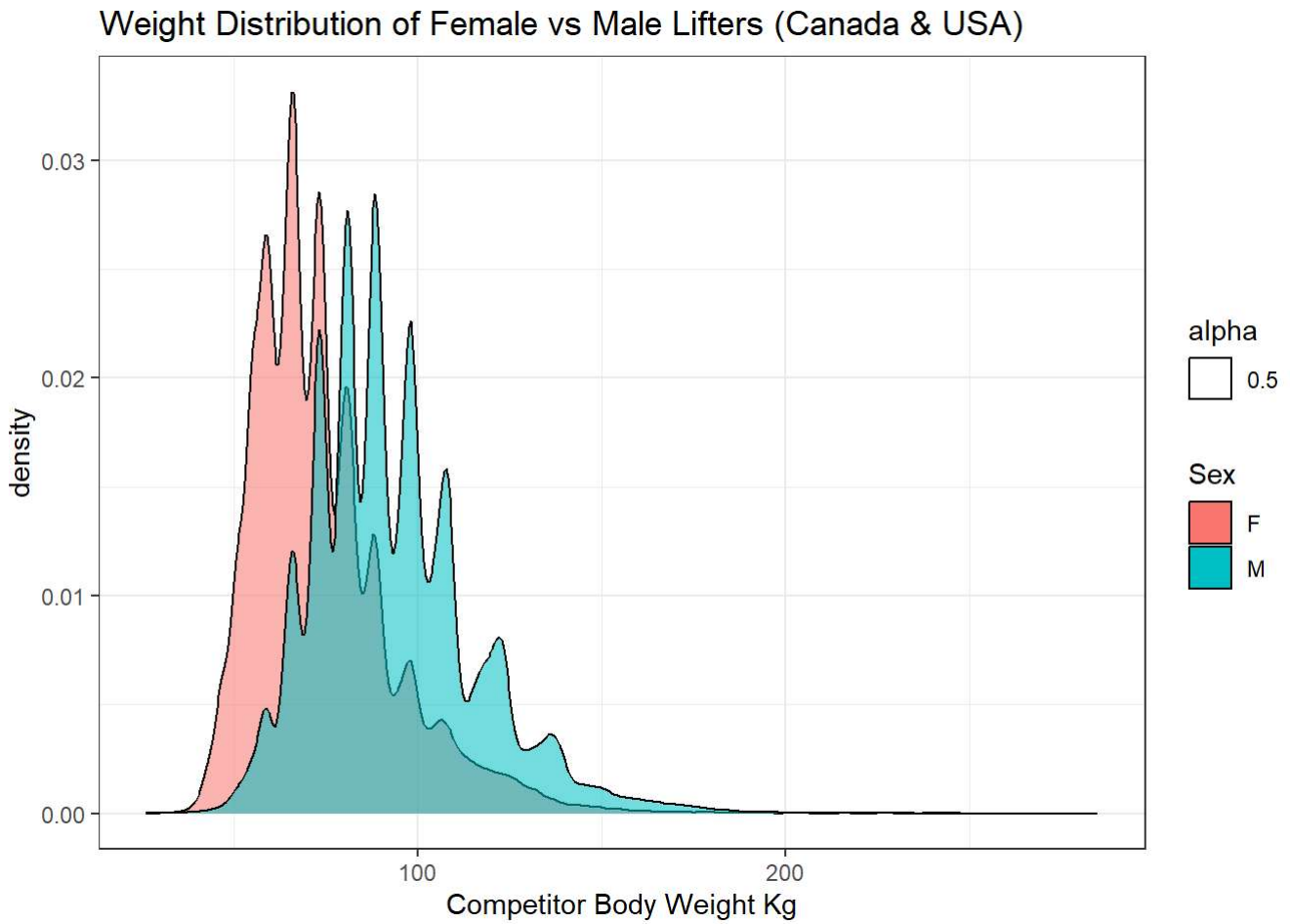
Expectedly, the majority of lifting competitors within Canada and the US are in their 20s. Canada. It's interesting to note the difference in the number of lifters in Canada compared to the US.Canada's largest age group is just over 300 competitors, while the US's larges age group has well over 10,000 competitors.

Weight Distribution

```
can_and_us_5 <- data %>% filter(MeetCountry %in% c("Canada", "USA"))

weight_data <- can_and_us_5 %>% drop_na(BodyweightKg) %>%
  mutate(BodyWeight = as.integer(as.numeric(gsub("\\+|\\s", "", BodyweightKg)))) %>%
  filter(Sex %in% c("M","F"))

weight_dist_plot <- ggplot(weight_data, aes(x=BodyWeight)) +
  geom_density(aes(fill=Sex, alpha=0.5)) +
  labs(title = "Weight Distribution of Female vs Male Lifters (Canada & USA)", x= "Competi
tor Body Weight Kg") + theme_bw()
weight_dist_plot
```




```
male_weights <- weight_data %>% filter(Sex=="M")
max(male_weights$BodyWeight)
```

```
## [1] 285
```

```
female_weights <- weight_data %>% filter(Sex=="F")
max(female_weights$BodyWeight)
```

```
## [1] 227
```

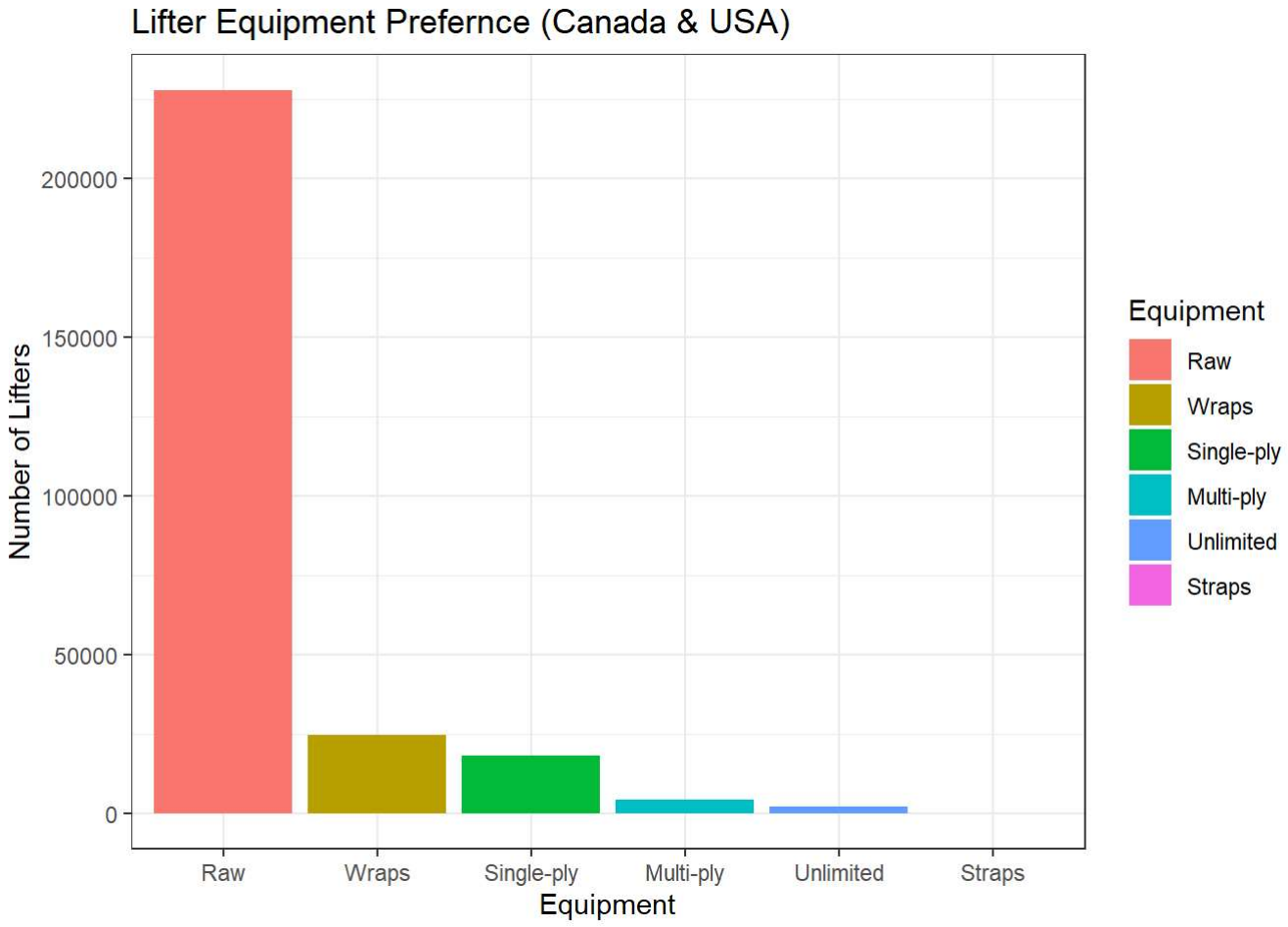
For Males, the most common body weight is close to 90 - 100 Kg, and around 60 - 70 Kg for Females. There are some interesting outliers though. Within the last 5 years there have been Male competitors weighing in at 285 Kg, and Female competitors weighing at 227 Kg.

Equipment Preference Distribution

```
assistive_equipment <- can_and_us_5 %>% drop_na(Equipment) %>% group_by(Equipment) %>% sum
marise(count=n())

assistive_equipment$Equipment <- reorder(assistive_equipment$Equipment, -assistive_equipme
nt$count)

assistive_equipment_plot <- ggplot(assistive_equipment, aes(x=Equipment, y=count, fill=Equ
ipment)) +
  geom_bar(stat = "identity") +
  labs(title = "Lifter Equipment Prefernce (Canada & USA)", x="Equipment", y="Number of Li
fters") + theme_bw()
assistive_equipment_plot
```



An overwhelming majority of lifters prefer to compete raw (without any assistive equipment).

LIFT ANALYSIS

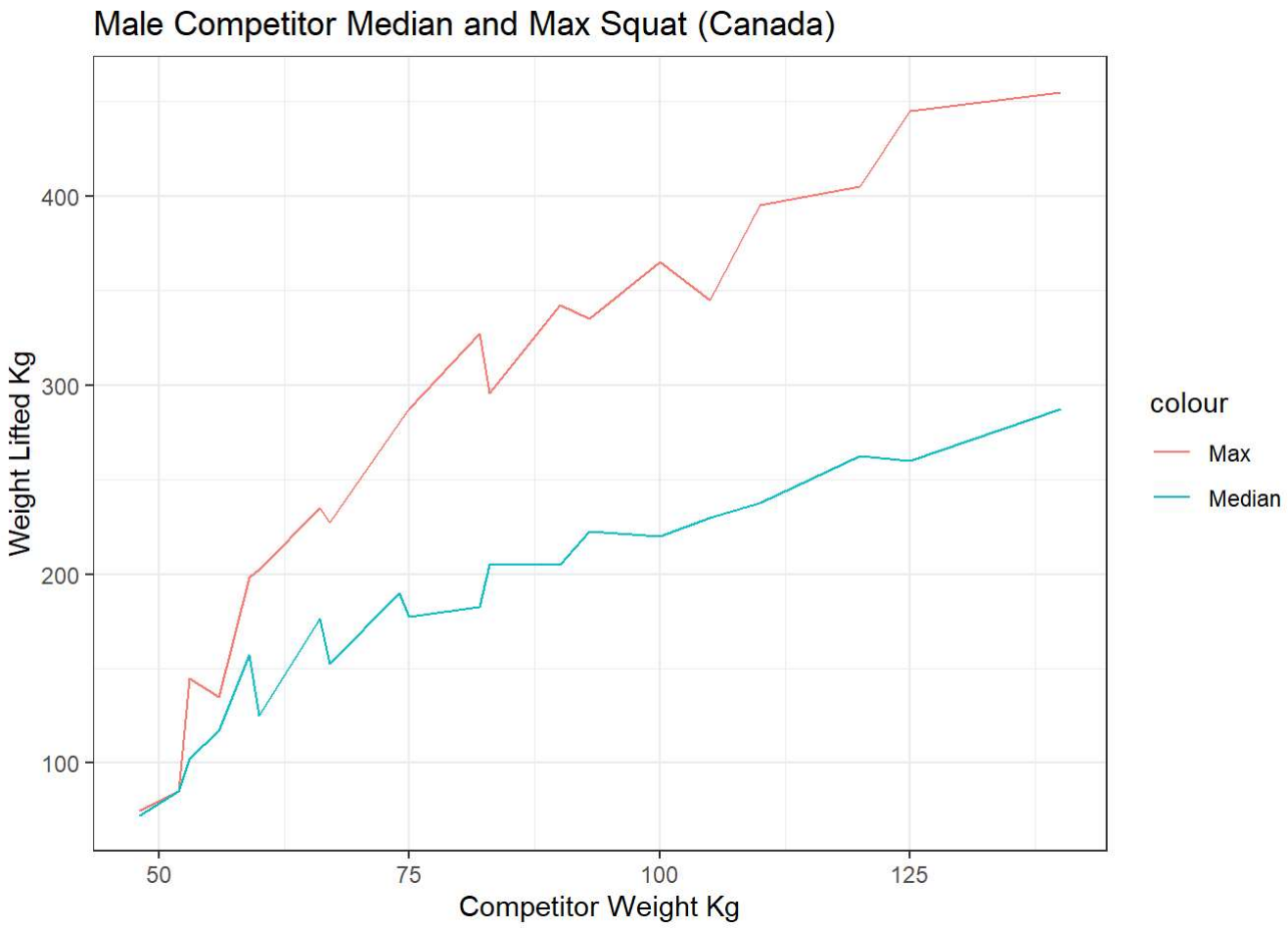
Let's see how my stats do among this data.

Squat

```
canada_meets = canada_meets %>% filter(!(WeightClassKg %in% c("","+","-"))) %>%
  mutate(WeightClass = as.integer(as.numeric(gsub("\\+|\\s", "", WeightClassKg))))

summary4 <- canada_meets %>% drop_na(Best3SquatKg) %>%
  filter(Sex=="M", Event %in% c("SBD","S")) %>% group_by(WeightClass) %>%
  summarise(best_lift = max(Best3SquatKg),
            median_lift = median(Best3SquatKg)) %>% arrange(desc(best_lift))

g4 <- ggplot(summary4, aes(x = WeightClass)) +
  geom_line(aes(y=best_lift, col="Max")) +
  geom_line(aes(y=median_lift, col="Median")) +
  labs(title = "Male Competitor Median and Max Squat (Canada)", x="Competitor Weight Kg",
y="Weight Lifted Kg") + theme_bw()
g4
```



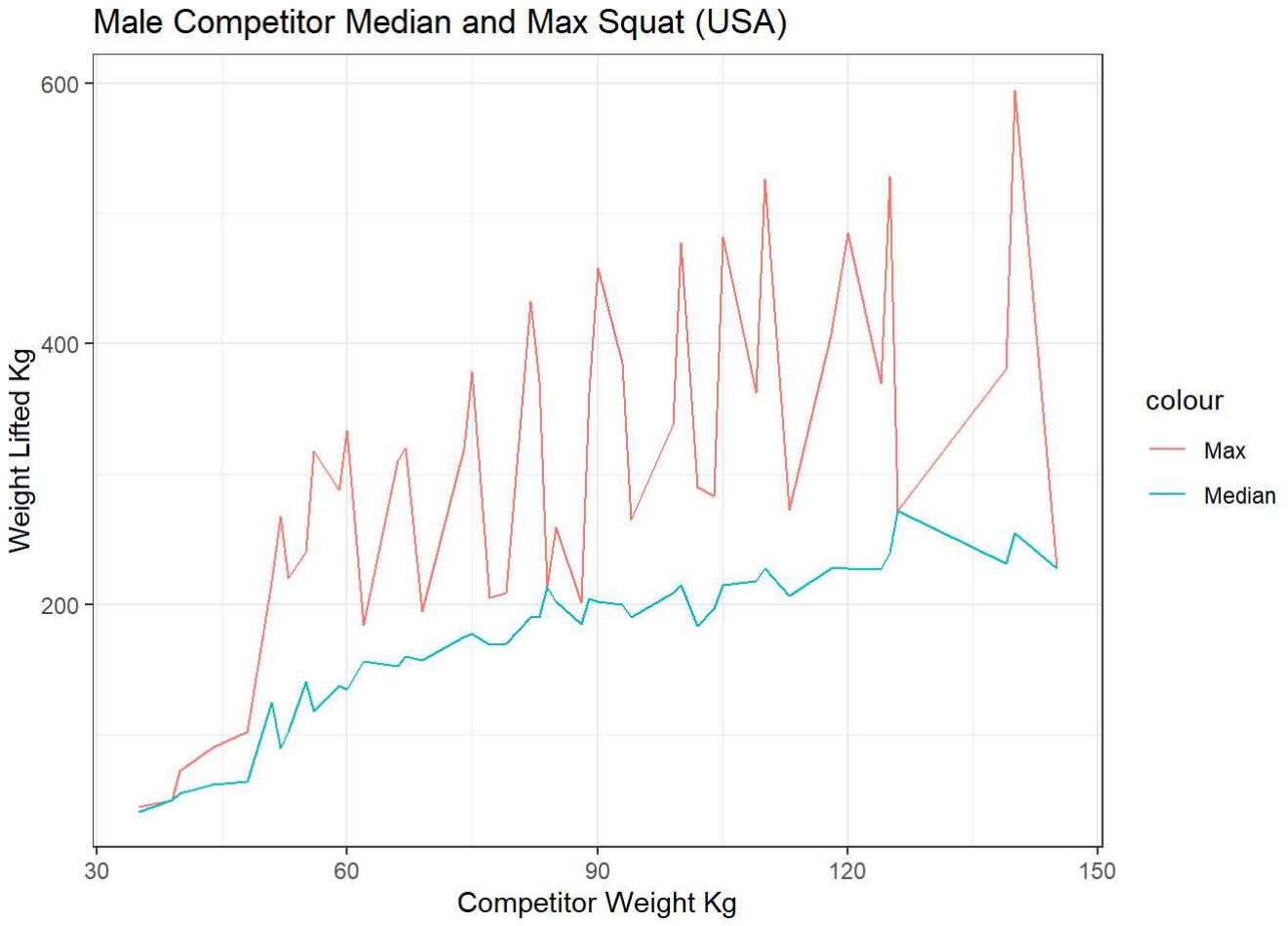

```
ecdf_fun <- ecdf(summary4$best_lift)
percentile_rank <- ecdf_fun(143.2) * 100
percentile_rank
```

```
## [1] 15
```

```
usa_meets = usa_meets %>% filter(!(WeightClassKg %in% c("","+","-"))) %>%
  mutate(WeightClass = as.integer(as.numeric(gsub("\\+|\\s", "", WeightClassKg))))

#Create a summary for male lifters
summary3 <- usa_meets %>% drop_na(Best3SquatKg) %>%
  filter(Sex=="M", Event %in% c("SBD","S")) %>% group_by(WeightClass) %>%
  summarise(best_lift = max(Best3SquatKg),
            median_lift = median(Best3SquatKg))

g3 <- ggplot(summary3, aes(x = WeightClass)) +
  geom_line(aes(y=best_lift, col="Max")) +
  geom_line(aes(y=median_lift, col="Median")) +
  labs(title = "Male Competitor Median and Max Squat (USA)", x="Competitor Weight Kg", y
="Weight Lifted Kg") + theme_bw()
g3
```




```
ecdf_fun <- ecdf(summary3$best_lift)
percentile_rank <- ecdf_fun(143.2) * 100
percentile_rank
```

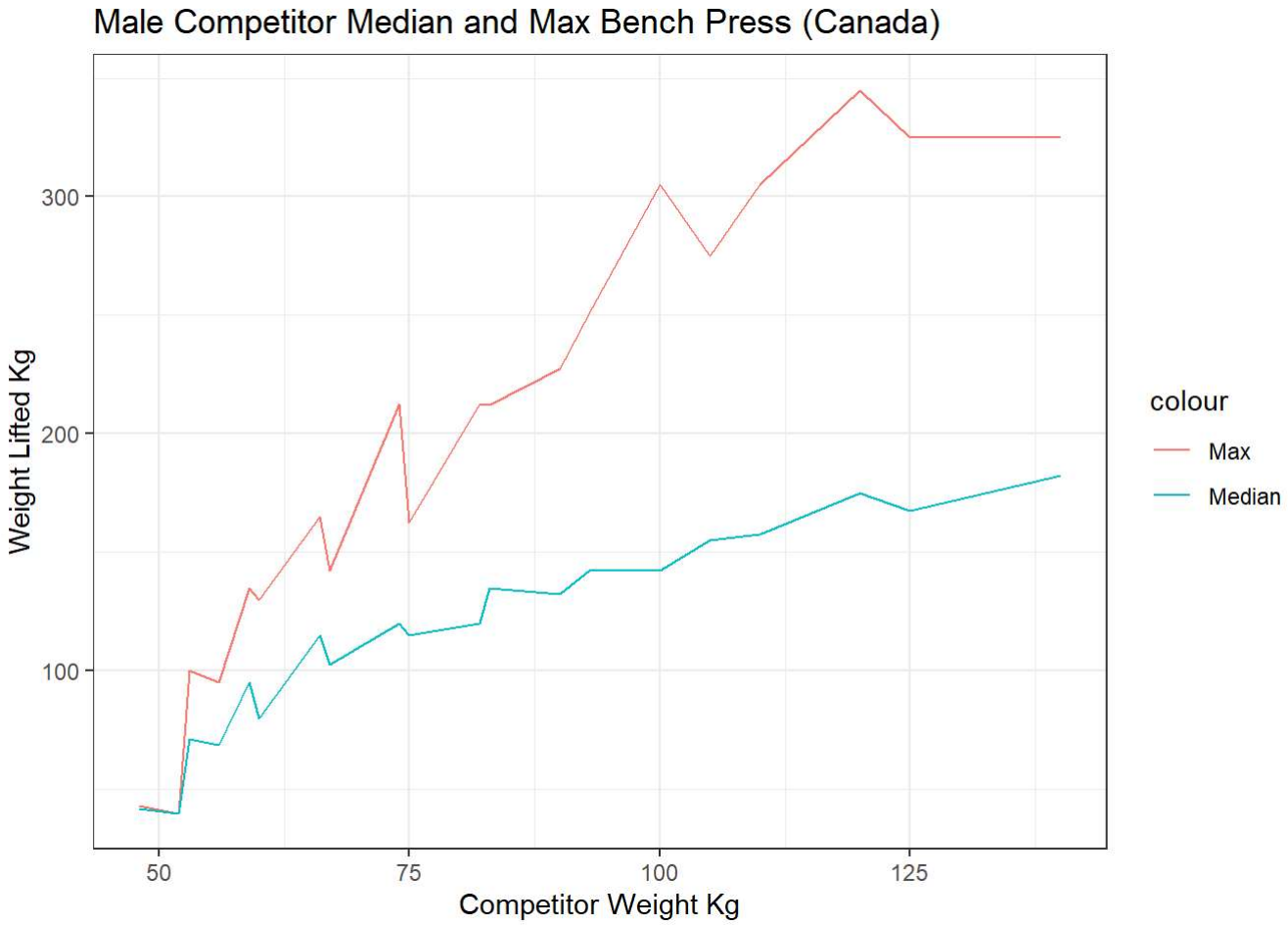
```
## [1] 11.11111
```

I have a body weight of about 100 Kg, and my max squat is 143.2 Kg. It looks like I am in the 15th percentile for Canadian lifters, and the 11th percentile for USA lifters. I clearly have a lot of work to do!

Bench Press

```
summary2 <- canada_meets %>% drop_na(Best3BenchKg) %>%
  filter(Sex=="M", Event %in% c("SBD","B")) %>% group_by(WeightClass) %>%
  summarise(best_lift = max(Best3BenchKg),
            median_lift = median(Best3BenchKg))

g1 <- ggplot(summary2, aes(x = WeightClass)) +
  geom_line(aes(y=best_lift, col="Max")) +
  geom_line(aes(y=median_lift, col="Median")) +
  labs(title = "Male Competitor Median and Max Bench Press (Canada)", x="Competitor Weight
Kg", y="Weight Lifted Kg") + theme_bw()
g1
```

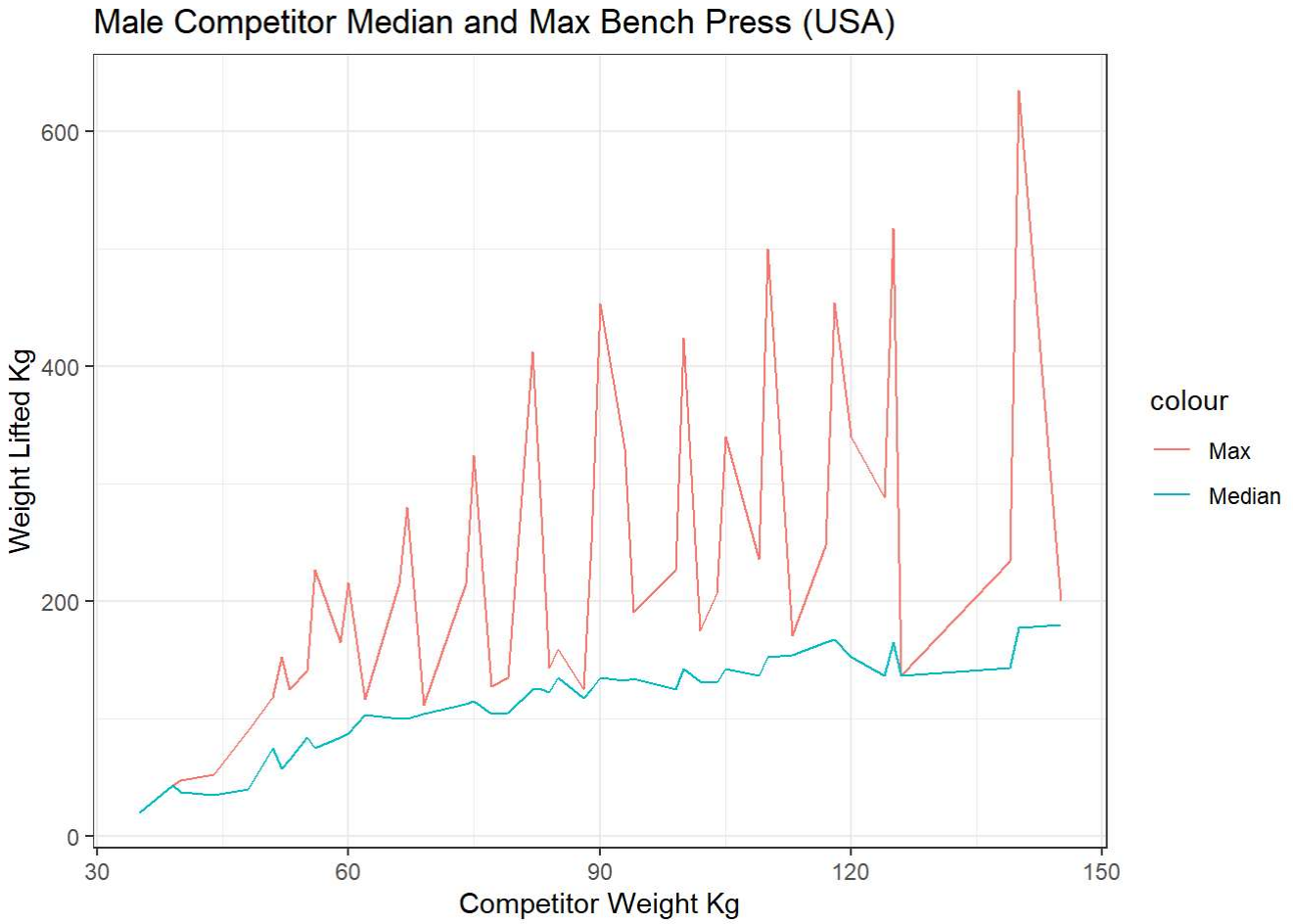



```
ecdf_fun <- ecdf(summary2$best_lift)
percentile_rank <- ecdf_fun(127.27) * 100
percentile_rank
```

```
## [1] 20
```

```
summary1 <- usa_meets %>% drop_na(Best3BenchKg) %>%
  filter(Sex=="M", Event %in% c("SBD","B")) %>% group_by(WeightClass) %>%
  summarise(best_lift = max(Best3BenchKg),
            median_lift = median(Best3BenchKg))

g <- ggplot(summary1, aes(x = WeightClass)) +
  geom_line(aes(y=best_lift, col="Max")) +
  geom_line(aes(y=median_lift, col="Median")) +
  labs(title = "Male Competitor Median and Max Bench Press (USA)", x="Competitor Weight K
g", y="Weight Lifted Kg") + theme_bw()
g
```




```
ecdf_fun <- ecdf(summary1$best_lift)
percentile_rank <- ecdf_fun(127.27) * 100
percentile_rank
```

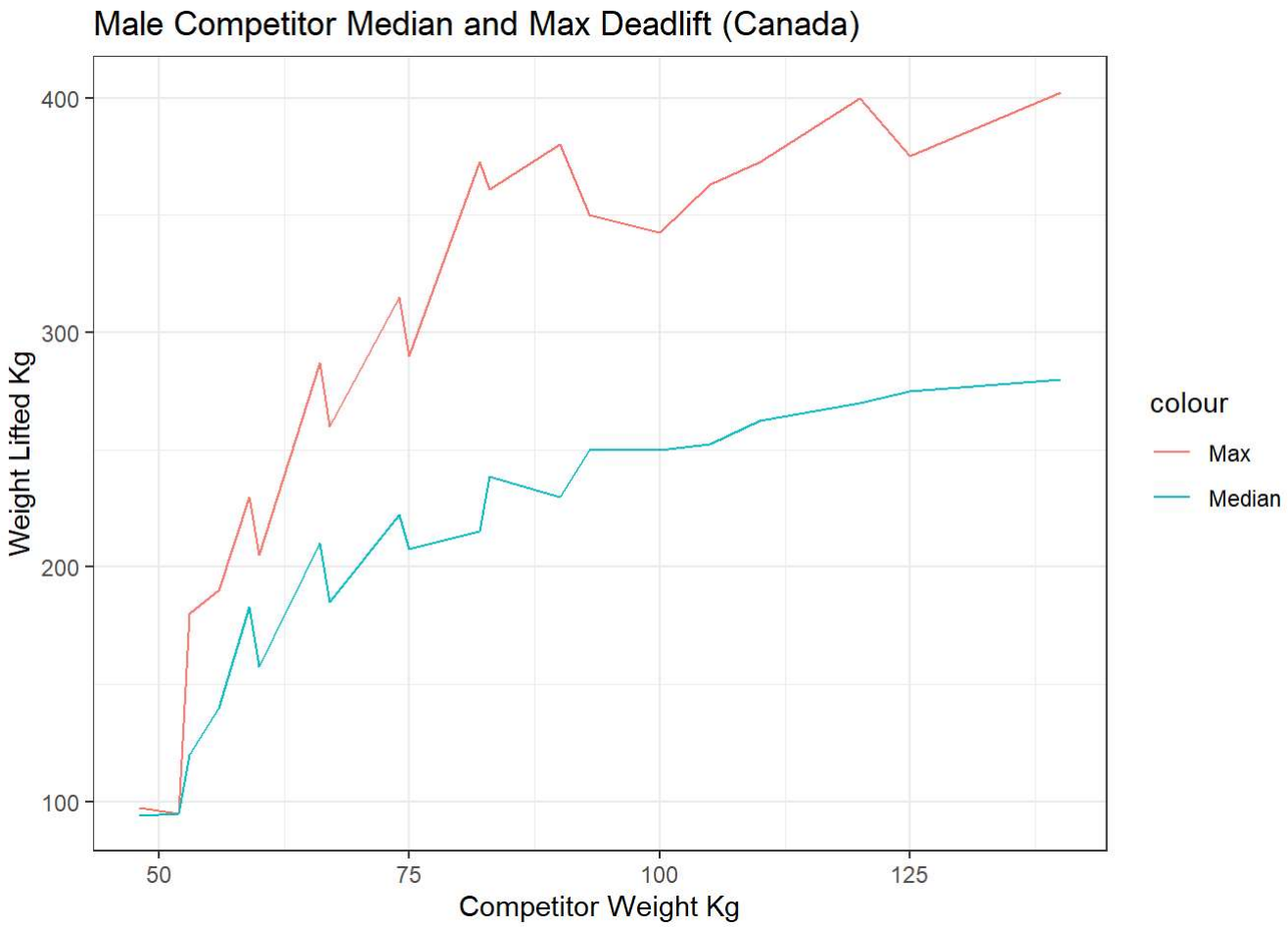
```
## [1] 23.91304
```

I have a max bench press of 127.27 Kg. This puts me in the 20th percentile among Canadian lifters and the 23rd percentile among USA lifters. A bit better than the squat. Hopefully I can one day enter the 30th percentiles for each.

Deadlift

```
summary6 <- canada_meets %>% drop_na(Best3DeadliftKg) %>%
  filter(Sex=="M", Event %in% c("SBD","D")) %>% group_by(WeightClass) %>%
  summarise(best_lift = max(Best3DeadliftKg),
            median_lift = median(Best3DeadliftKg))

g6 <- ggplot(summary6, aes(x = WeightClass)) +
  geom_line(aes(y=best_lift, col="Max")) +
  geom_line(aes(y=median_lift, col="Median")) +
  labs(title = "Male Competitor Median and Max Deadlift (Canada)", x="Competitor Weight K
g", y="Weight Lifted Kg") + theme_bw()
g6
```

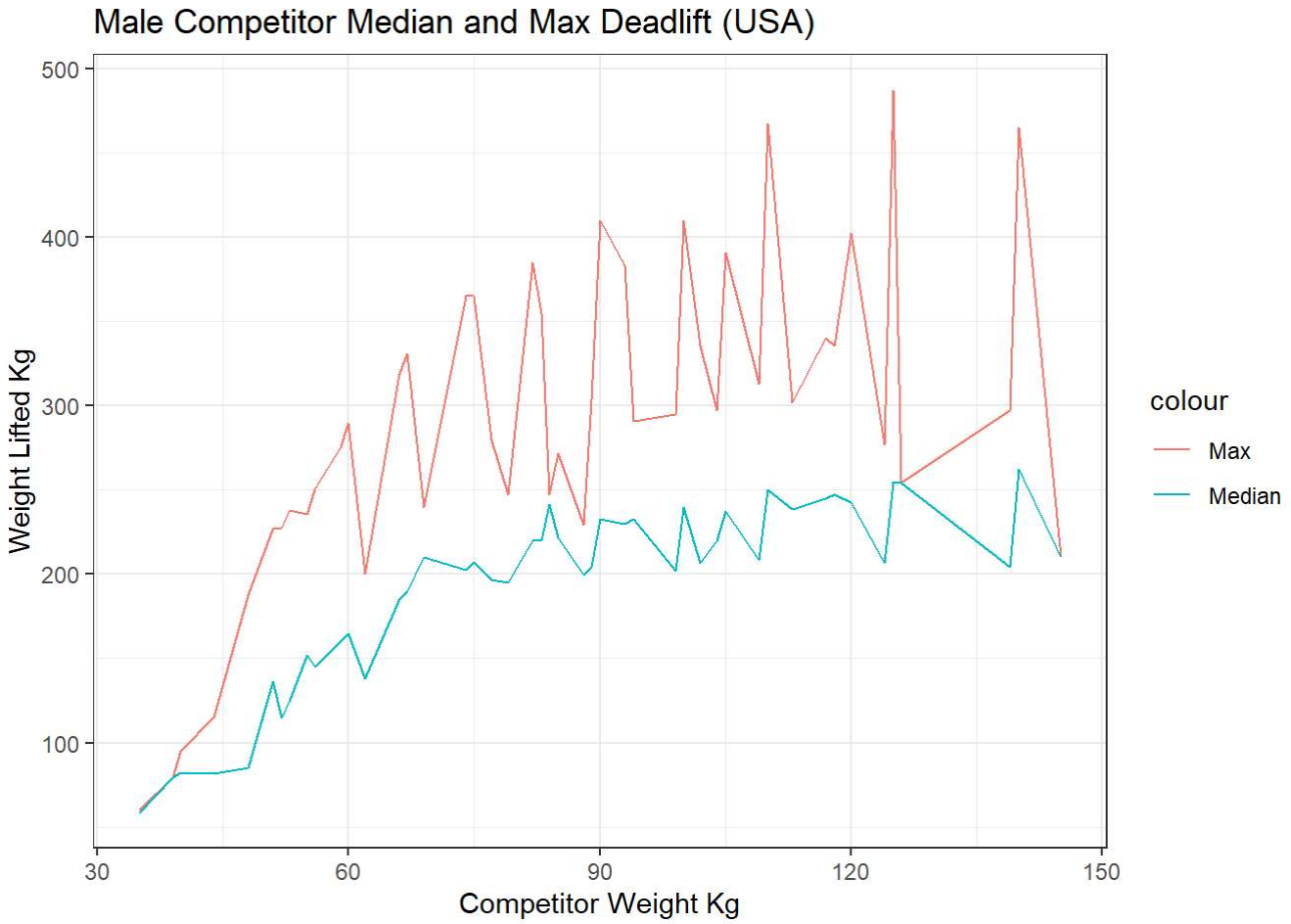



```
ecdf_fun <- ecdf(summary6$best_lift)
percentile_rank <- ecdf_fun(215.9) * 100
percentile_rank
```

```
## [1] 25
```

```
summary5 <- usa_meets %>% drop_na(Best3DeadliftKg) %>%
  filter(Sex=="M", Event %in% c("SBD","D")) %>% group_by(WeightClass) %>%
  summarise(best_lift = max(Best3DeadliftKg),
            median_lift = median(Best3DeadliftKg))

g5 <- ggplot(summary5, aes(x = WeightClass)) +
  geom_line(aes(y=best_lift, col="Max")) +
  geom_line(aes(y=median_lift, col="Median")) +
  labs(title = "Male Competitor Median and Max Deadlift (USA)", x="Competitor Weight Kg",
y="Weight Lifted Kg") + theme_bw()
g5
```




```
ecdf_fun <- ecdf(summary5$best_lift)
percentile_rank <- ecdf_fun(215.9) * 100
percentile_rank
```

```
## [1] 15.21739
```

Lastly, my max deadlift is 215.9 Kg. This will put me in the among 25th percentile of Canadian lifters, and only the 15th percentile among USA lifters. So far, this is my best lift among Canadian lifters and I am proud of this.

Aside from my current standings, another piece of information we can take away is that USA lifters are generally better competitors than the Canadians in terms of their best lifts. This is not too surprising considering that the difference in the number of competitors within both countries.