

# Propensity Score Diagnostics

Lucy D'Agostino McGowan

Wake Forest University

2020-07-29 (updated: 2020-12-01)

# Checking balance

## Love plots (Standardized Mean Difference)

## ECDF plots

# Standardized Mean Difference (SMD)

$$d = \frac{\bar{x}_{treatment} - \bar{x}_{control}}{\sqrt{\frac{s_{treatment}^2 + s_{control}^2}{2}}}$$

# SMD in R

## 1 Create a "design object" to incorporate the weights

```
library(survey)

svy_des <- svydesign(
  ids = ~ 1,
  data = df,
  weights = ~ wts
)
```

# SMD in R

## 2 Calculate the unweighted standardized mean differences

```
library(tableone)
library(tidyverse)

smd_table_unweighted <- CreateTableOne(
  vars = c("confounder_1", "confounder_1", ...),
  strata = "exposure",
  data = df,
  test = FALSE)
```

# SMD in R

## 3 Calculate the weighted standardized mean differences

```
smd_table <- svyCreateTableOne(  
  vars = c("confounder_1", "confounder_1", ...),  
  strata = "exposure",  
  data = svy_des,  
  test = FALSE)
```

# SMD in R

## 3 Calculate the weighted standardized mean differences

```
smd_table <- svyCreateTableOne(  
  vars = c("confounder_1", "confounder_1", ...),  
  strata = "exposure",  
  data = svy_des,  
  test = FALSE)
```

# SMD in R

## 4 Stick these together in a data frame

```
plot_df <- data.frame(  
  var = rownames(ExtractSmd(smd_table)),  
  Unadjusted = as.numeric(ExtractSmd(smd_table_unweighted)),  
  Weighted = as.numeric(ExtractSmd(smd_table))) %>%  
  pivot_longer(-var, names_to = "Method", values_to = "SMD")
```



# SMD in R

## 4 Stick these together in a data frame

```
plot_df <- data.frame(  
  var = rownames(ExtractSmd(smd_table)),  
  Unadjusted = as.numeric(ExtractSmd(smd_table_unweighted)),  
  Weighted = as.numeric(ExtractSmd(smd_table))) %>%  
  pivot_longer(-var, names_to = "Method", values_to = "SMD")  
  
rownames(ExtractSMD(smd_table))  
#> [1] "confounder_1" "confounder_2"
```

# SMD in R

## 4 Stick these together in a data frame

```
plot_df <- data.frame(  
  var = rownames(ExtractSmd(smd_table)),  
  Unadjusted = as.numeric(ExtractSmd(smd_table_unweighted)),  
  Weighted = as.numeric(ExtractSmd(smd_table)) %>%  
  pivot_longer(-var, names_to = "Method", values_to = "SMD")  
  
as.numeric(ExtractSmd(smd_table_unweighted))  
#> [1] 0.160 0.177
```

# SMD in R

## 4 Stick these together in a data frame

```
plot_df <- data.frame(  
  var = rownames(ExtractSmd(smd_table)),  
  Unadjusted = as.numeric(ExtractSmd(smd_table_unweighted)),  
  Weighted = as.numeric(ExtractSmd(smd_table)) %>%  
  pivot_longer(-var, names_to = "Method", values_to = "SMD")  
  
as.numeric(ExtractSmd(smd_table))  
#> [1] 0.002 0.007
```

# SMD in R

## 4 Stick these together in a data frame

```
plot_df <- data.frame(  
  var = rownames(ExtractSmd(smd_table)),  
  Unadjusted = as.numeric(ExtractSmd(smd_table_unweighted)),  
  Weighted = as.numeric(ExtractSmd(smd_table))) %>%  
  pivot_longer(-var, names_to = "Method", values_to = "SMD")
```

# SMD in R

## 5 Plot them! (in a Love plot!)

```
ggplot(data = plot_df,  
       mapping = aes(x = var, y = SMD, group = Method, color = Method),  
       geom_line() +  
       geom_point() +  
       geom_hline(yintercept = 0.1, color = "black", size = 0.1) +  
       coord_flip())
```

# SMD in R

## 5 Plot them! (in a Love plot!)

```
ggplot(data = plot_df,  
       mapping = aes(x = var, y = SMD, group = Method, color = Method)) +  
  geom_line() +  
  geom_point() +  
  geom_hline(yintercept = 0.1, color = "black", size = 0.1) +  
  coord_flip()
```

# SMD in R

## 5 Plot them! (in a Love plot!)

```
ggplot(data = plot_df,  
       mapping = aes(x = var, y = SMD, group = Method, color = Method)) +  
  geom_line() +  
  geom_point() +  
  geom_hline(yintercept = 0.1, color = "black", size = 0.1) +  
  coord_flip()
```

# SMD in R

## 5 Plot them! (in a Love plot!)

```
ggplot(data = plot_df,  
       mapping = aes(x = var, y = SMD, group = Method, color = Method),  
       geom_line() +  
       geom_point() +  
       geom_hline(yintercept = 0.1, color = "black", size = 0.1) +  
       coord_flip())
```



# SMD in R

## 5 Plot them! (in a Love plot!)

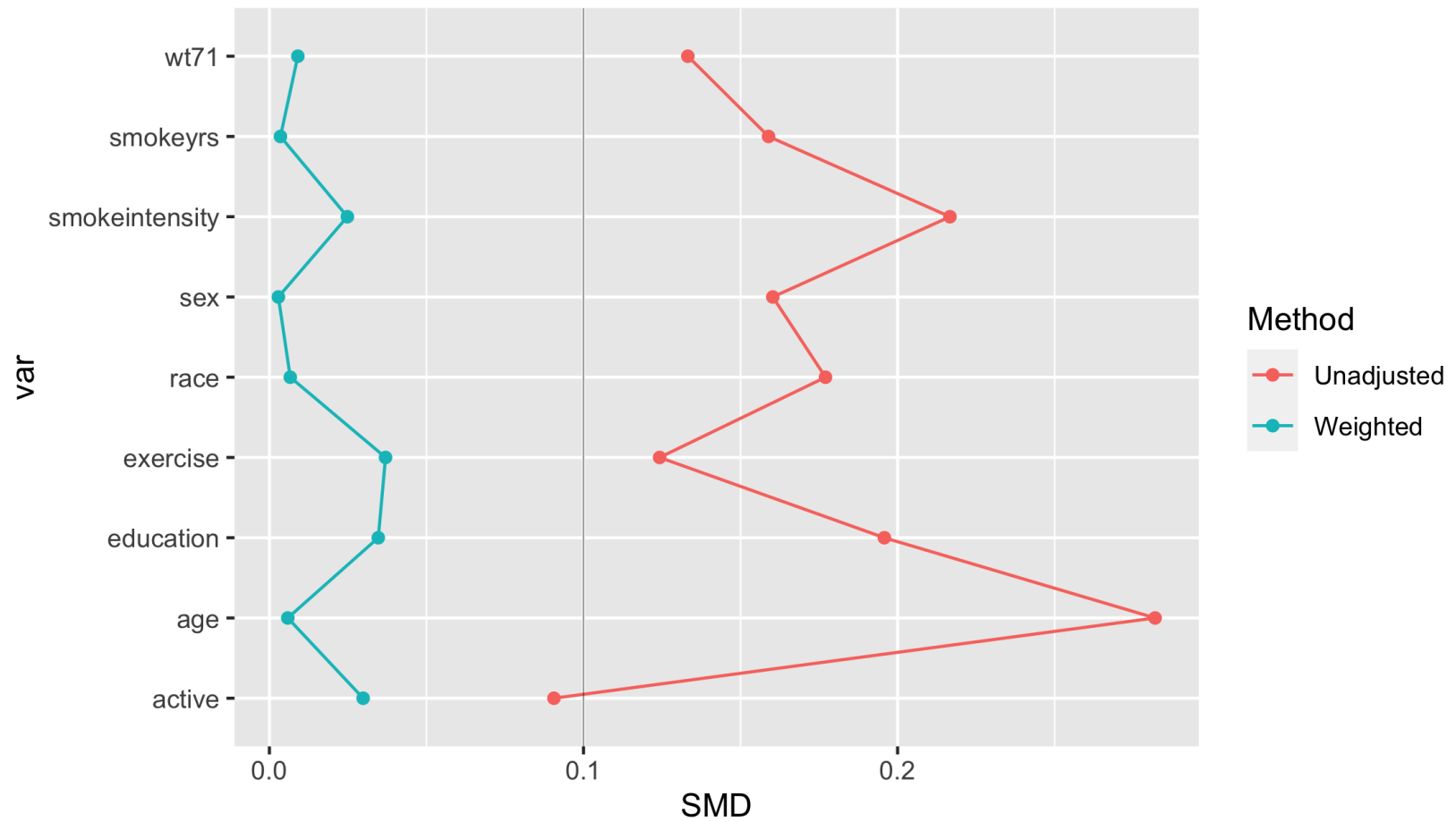
```
ggplot(data = plot_df,  
       mapping = aes(x = var, y = SMD, group = Method, color = Method),  
       geom_line() +  
       geom_point() +  
       geom_hline(yintercept = 0.1, color = "black", size = 0.1) +  
       coord_flip())
```

# SMD in R

## 5 Plot them! (in a Love plot!)

```
ggplot(data = plot_df, mapping = aes(x = var, y = SMD, group = Method)) +  
  geom_line() +  
  geom_point() +  
  geom_hline(yintercept = 0.1, color = "black", size = 0.1) +  
  coord_flip()
```

# Love plot



## Your turn 1

- 1 Create a Love Plot for the propensity score weighting you created in the previous exercise**

07:00

# ECDF

For continuous variables, it can be helpful to look at the **whole** distribution pre and post-weighting rather than a single summary measure

# Unweighted ECDF

```
ggplot(df, aes(x = wt71, group = qsmk, color = factor(qsmk))) +  
  stat_ecdf() +  
  scale_color_manual("Quit smoking", values = c("#5154B8", "#5DB854"),  
                    labels = c("Yes", "No")) +  
  xlab("Weight in Kg in 1971") +  
  ylab("Proportion <= x")
```

# Unweighted ECDF

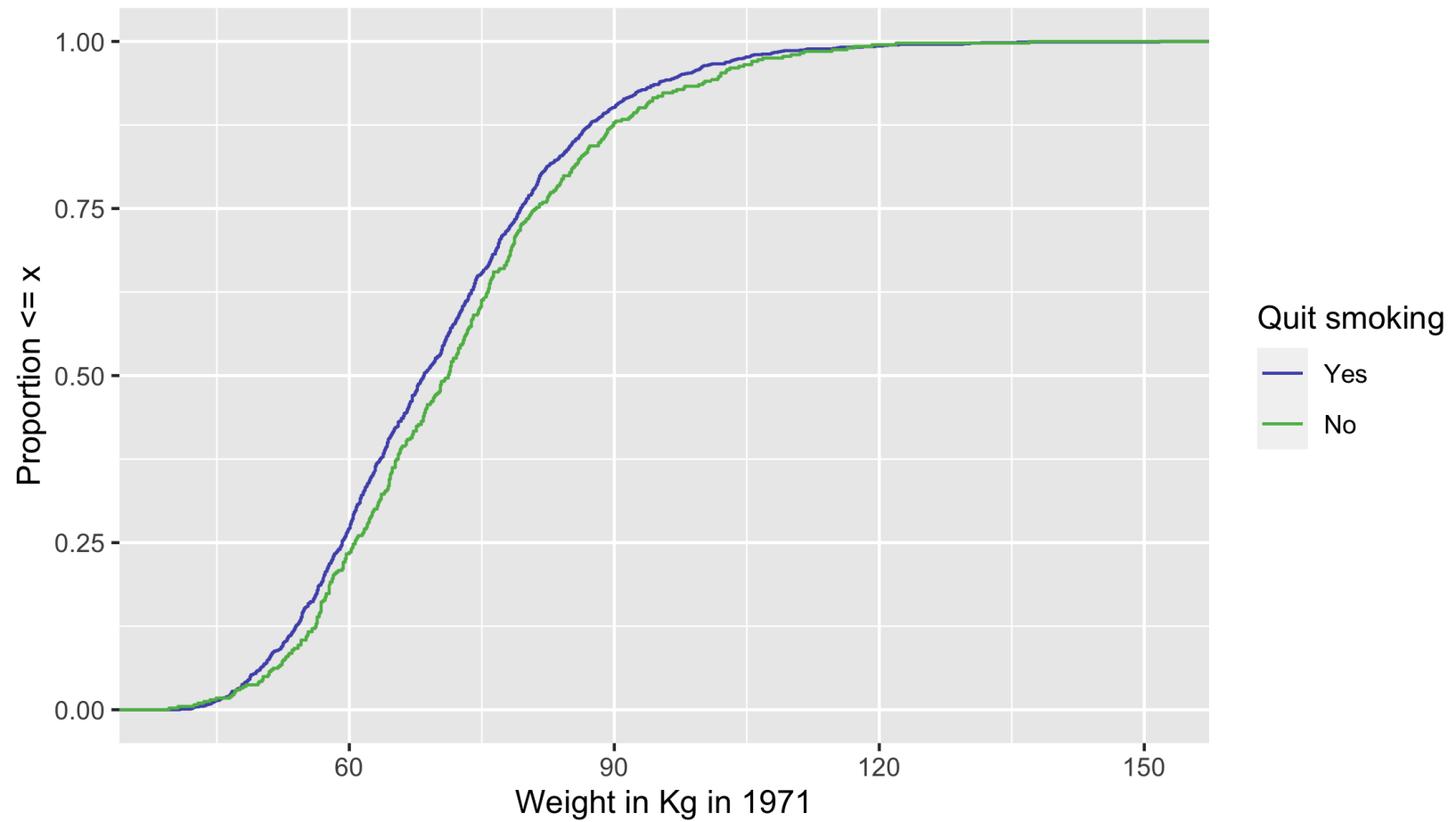
```
ggplot(df, aes(x = wt71, group = qsmk, color = factor(qsmk))) +  
  stat_ecdf() +  
  scale_color_manual("Quit smoking", values = c("#5154B8", "#5DB854"),  
                    labels = c("Yes", "No")) +  
  xlab("Weight in Kg in 1971") +  
  ylab("Proportion <= x")
```

# Unweighted ECDF

```
ggplot(df, aes(x = wt71, group = qsmk, color = factor(qsmk))) +  
  stat_ecdf() +  
  scale_color_manual("Quit smoking", values = c("#5154B8", "#5DB854"),  
                    labels = c("Yes", "No")) +  
  xlab("Weight in Kg in 1971") +  
  ylab("Proportion <= x")
```



# Unweighted ECDF



# Weighted ECDF

```
ecdf_1 <- df %>%  
  filter(qsmk == 1) %>%  
  arrange(wt71) %>%  
  mutate(cum_pct = cumsum(w_ate) / sum(w_ate))  
  
ecdf_0 <- df %>%  
  filter(qsmk == 0) %>%  
  arrange(wt71) %>%  
  mutate(cum_pct = cumsum(w_ate) / sum(w_ate))  
  
ggplot(ecdf_1, aes(x = wt71, y = cum_pct)) +  
  geom_line( color = "#5DB854") +  
  geom_line(data = ecdf_0, aes(x = wt71, y = cum_pct), color = "#515.5") +  
  xlab("Weight in Kg in 1971") +  
  ylab("Proportion <= x")
```

# Weighted ECDF

```
ecdf_1 <- df %>%  
  filter(qsmk == 1) %>%  
  arrange(wt71) %>%  
  mutate(cum_pct = cumsum(w_ate) / sum(w_ate))  
  
ecdf_0 <- df %>%  
  filter(qsmk == 0) %>%  
  arrange(wt71) %>%  
  mutate(cum_pct = cumsum(w_ate) / sum(w_ate))  
  
ggplot(ecdf_1, aes(x = wt71, y = cum_pct)) +  
  geom_line( color = "#5DB854") +  
  geom_line(data = ecdf_0, aes(x = wt71, y = cum_pct), color = "#515151") +  
  xlab("Weight in Kg in 1971") +  
  ylab("Proportion <= x")
```

# Weighted ECDF

```
ecdf_1 <- df %>%  
  filter(qsmk == 1) %>%  
  arrange(wt71) %>%  
  mutate(cum_pct = cumsum(w_ate) / sum(w_ate))  
  
ecdf_0 <- df %>%  
  filter(qsmk == 0) %>%  
  arrange(wt71) %>%  
  mutate(cum_pct = cumsum(w_ate) / sum(w_ate))  
  
ggplot(ecdf_1, aes(x = wt71, y = cum_pct)) +  
  geom_line( color = "#5DB854") +  
  geom_line(data = ecdf_0, aes(x = wt71, y = cum_pct), color = "#515151") +  
  xlab("Weight in Kg in 1971") +  
  ylab("Proportion <= x")
```

# Weighted ECDF

```
ecdf_1 <- df %>%  
  filter(qsmk == 1) %>%  
  arrange(wt71) %>%  
  mutate(cum_pct = cumsum(w_ate) / sum(w_ate))  
  
ecdf_0 <- df %>%  
  filter(qsmk == 0) %>%  
  arrange(wt71) %>%  
  mutate(cum_pct = cumsum(w_ate) / sum(w_ate))  
  
ggplot(ecdf_1, aes(x = wt71, y = cum_pct)) +  
  geom_line( color = "#5DB854") +  
  geom_line(data = ecdf_0, aes(x = wt71, y = cum_pct), color = "#515.  
  xlab("Weight in Kg in 1971") +  
  ylab("Proportion <= x")
```

# Weighted ECDF

```
ecdf_1 <- df %>%  
  filter(qsmk == 1) %>%  
  arrange(wt71) %>%  
  mutate(cum_pct = cumsum(w_ate) / sum(w_ate))
```

```
ecdf_0 <- df %>%  
  filter(qsmk == 0) %>%  
  arrange(wt71) %>%  
  mutate(cum_pct = cumsum(w_ate) / sum(w_ate))
```

```
ggplot(ecdf_1, aes(x = wt71, y = cum_pct)) +  
  geom_line( color = "#5DB854") +  
  geom_line(data = ecdf_0, aes(x = wt71, y = cum_pct), color = "#515.  
  xlab("Weight in Kg in 1971") +  
  ylab("Proportion <= x")
```

# Weighted ECDF

```
ecdf_1 <- df %>%  
  filter(qsmk == 1) %>%  
  arrange(wt71) %>%  
  mutate(cum_pct = cumsum(w_ate) / sum(w_ate))  
  
ecdf_0 <- df %>%  
  filter(qsmk == 0) %>%  
  arrange(wt71) %>%  
  mutate(cum_pct = cumsum(w_ate) / sum(w_ate))  
  
ggplot(ecdf_1, aes(x = wt71, y = cum_pct)) +  
  geom_line( color = "#5DB854") +  
  geom_line(data = ecdf_0, aes(x = wt71, y = cum_pct), color = "#515151") +  
  xlab("Weight in Kg in 1971") +  
  ylab("Proportion <= x")
```

# Weighted ECDF

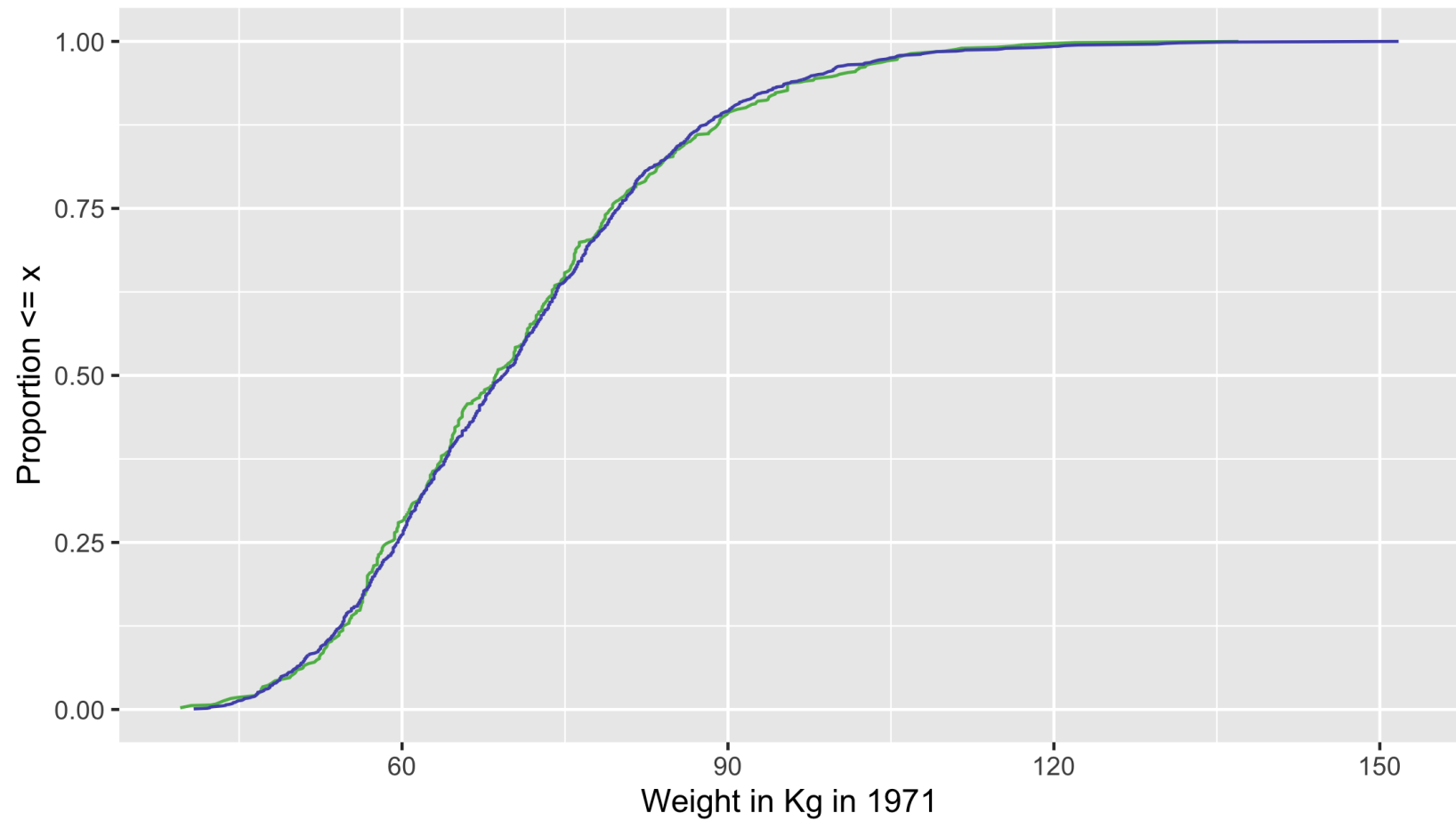
```
ecdf_1 <- df %>%  
  filter(qsmk == 1) %>%  
  arrange(wt71) %>%  
  mutate(cum_pct = cumsum(w_ate) / sum(w_ate))  
  
ecdf_0 <- df %>%  
  filter(qsmk == 0) %>%  
  arrange(wt71) %>%  
  mutate(cum_pct = cumsum(w_ate) / sum(w_ate))  
  
ggplot(ecdf_1, aes(x = wt71, y = cum_pct)) +  
  geom_line( color = "#5DB854") +  
  geom_line(data = ecdf_0, aes(x = wt71, y = cum_pct), color = "#515.5")  
  xlab("Weight in Kg in 1971") +  
  ylab("Proportion <= x")
```



# Weighted ECDF

```
ecdf_1 <- df %>%  
  filter(qsmk == 1) %>%  
  arrange(wt71) %>%  
  mutate(cum_pct = cumsum(w_ate) / sum(w_ate))  
  
ecdf_0 <- df %>%  
  filter(qsmk == 0) %>%  
  arrange(wt71) %>%  
  mutate(cum_pct = cumsum(w_ate) / sum(w_ate))  
  
ggplot(ecdf_1, aes(x = wt71, y = cum_pct)) +  
  geom_line( color = "#5DB854") +  
  geom_line(data = ecdf_0, aes(x = wt71, y = cum_pct), color = "#515151") +  
  xlab("Weight in Kg in 1971") +  
  ylab("Proportion <= x")
```

# Weighted ECDF



## Your turn 2

- 1 Create an unweighted ECDF examining the smokeyrs confounder for those that quit smoking and those that did not**
- 2 Create a weighted ECDF examining the smokeyrs confounder**

07:00