

Propensity Score Diagnostics

Lucy D'Agostino McGowan
Wake Forest University

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



Calculate standardized mean differences

```
1 library(halfmoon)
2 library(tidyverse)
3
4 smds <- tidy_smd(
5   df,
6   .vars = c(confounder_1, confounder_2, ...),
7   .group = exposure,
8   .wts = wts # optional,
9   make_dummy_vars = TRUE # optional
10 )
```

Calculating SMDs

```
1 vars <- c(  
2   "sex", "race", "age", "education",  
3   "smokeintensity", "smokeyr",  
4   "exercise", "active", "wt71"  
5 )  
6  
7 smds <- tidy_smd(  
8   nhfs_complete_wts,  
9   .vars = all_of(vars),  
10  .group = qsmk,  
11  .wts = w_ate,  
12  make_dummy_vars = TRUE  
13 )  
14  
15 smds
```

Calculating SMDs

```
# A tibble: 28 × 4
```

	variable <chr>	method <chr>	qsmk <chr>	smd <dbl>
1	sex1	observed	1	0.160
2	race1	observed	1	0.177
3	age	observed	1	-0.282
4	education2	observed	1	0.112
5	education3	observed	1	0.0472
6	education4	observed	1	0.0270
7	education5	observed	1	-0.166
8	smokeintensity	observed	1	0.217
9	smokeys	observed	1	-0.159
10	exercise1	observed	1	-0.0398

```
"# A tibble: 18 × 4"
```

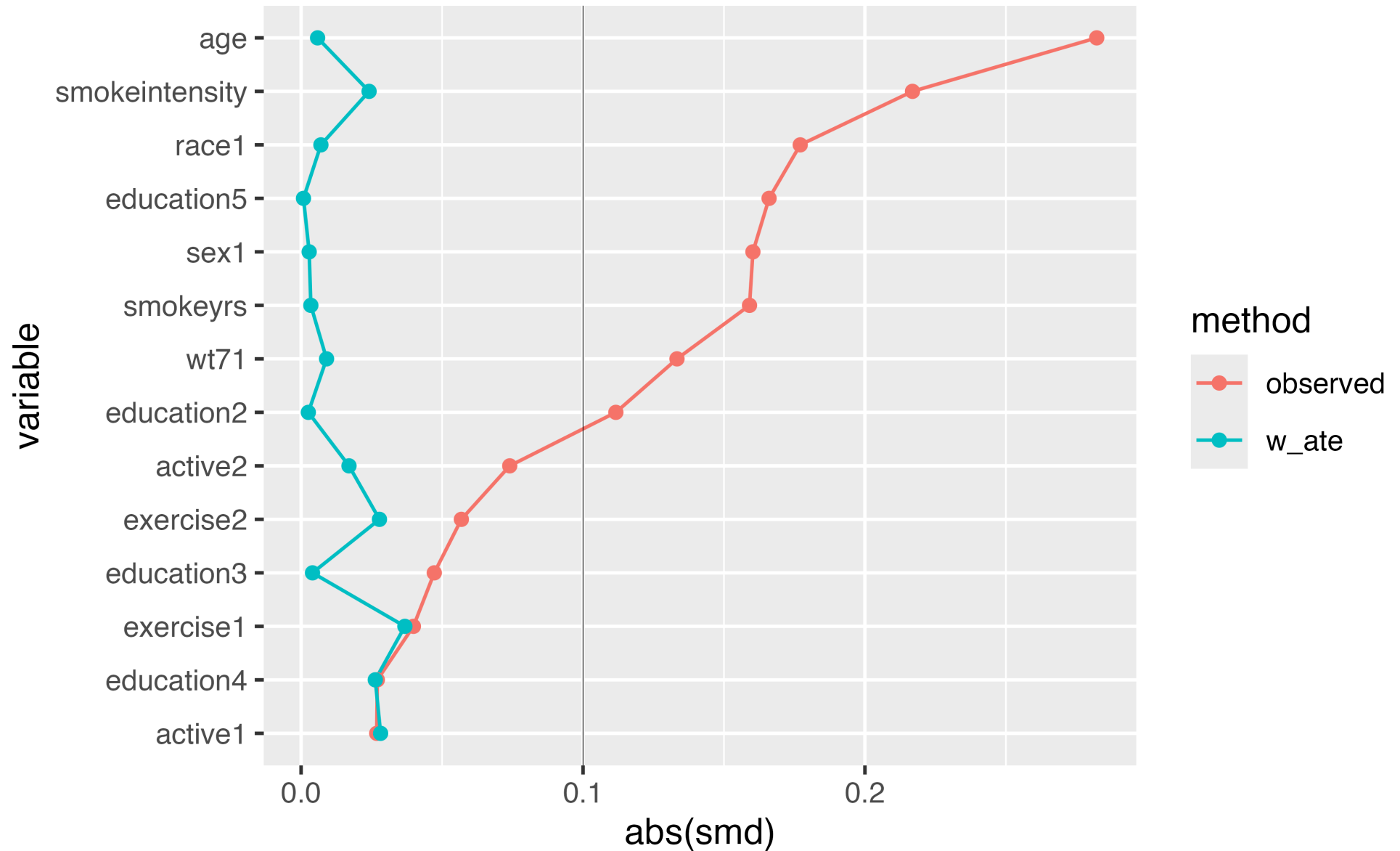
Plotting SMDs



Plot them! (in a Love plot!)

```
1 ggplot(  
2   data = smds,  
3   aes(  
4     x = abs(smd),  
5     y = variable,  
6     group = method,  
7     color = method  
8   )  
9 ) +  
10  geom_love()
```

Love plot



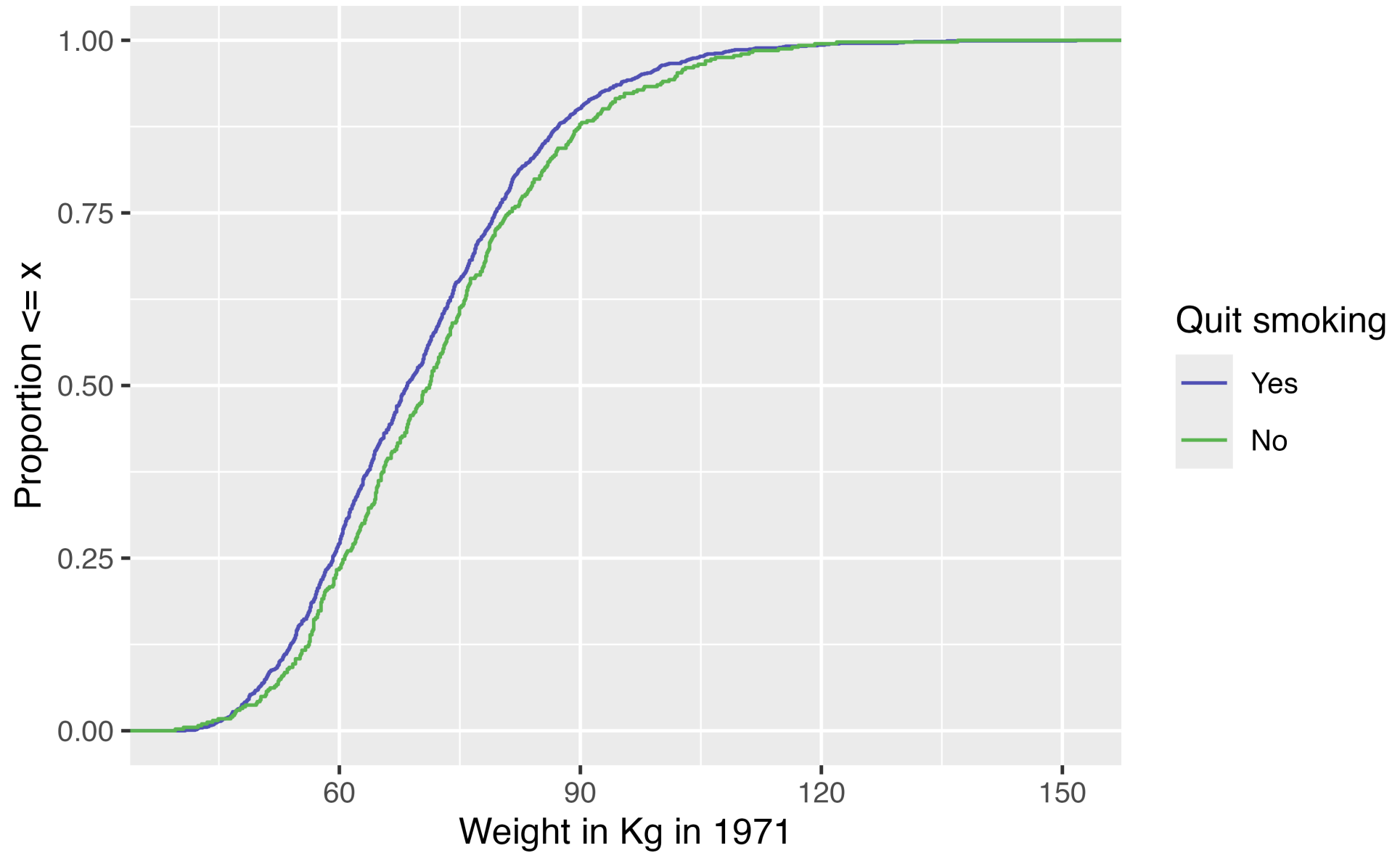
Your turn 1

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

ECDF

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

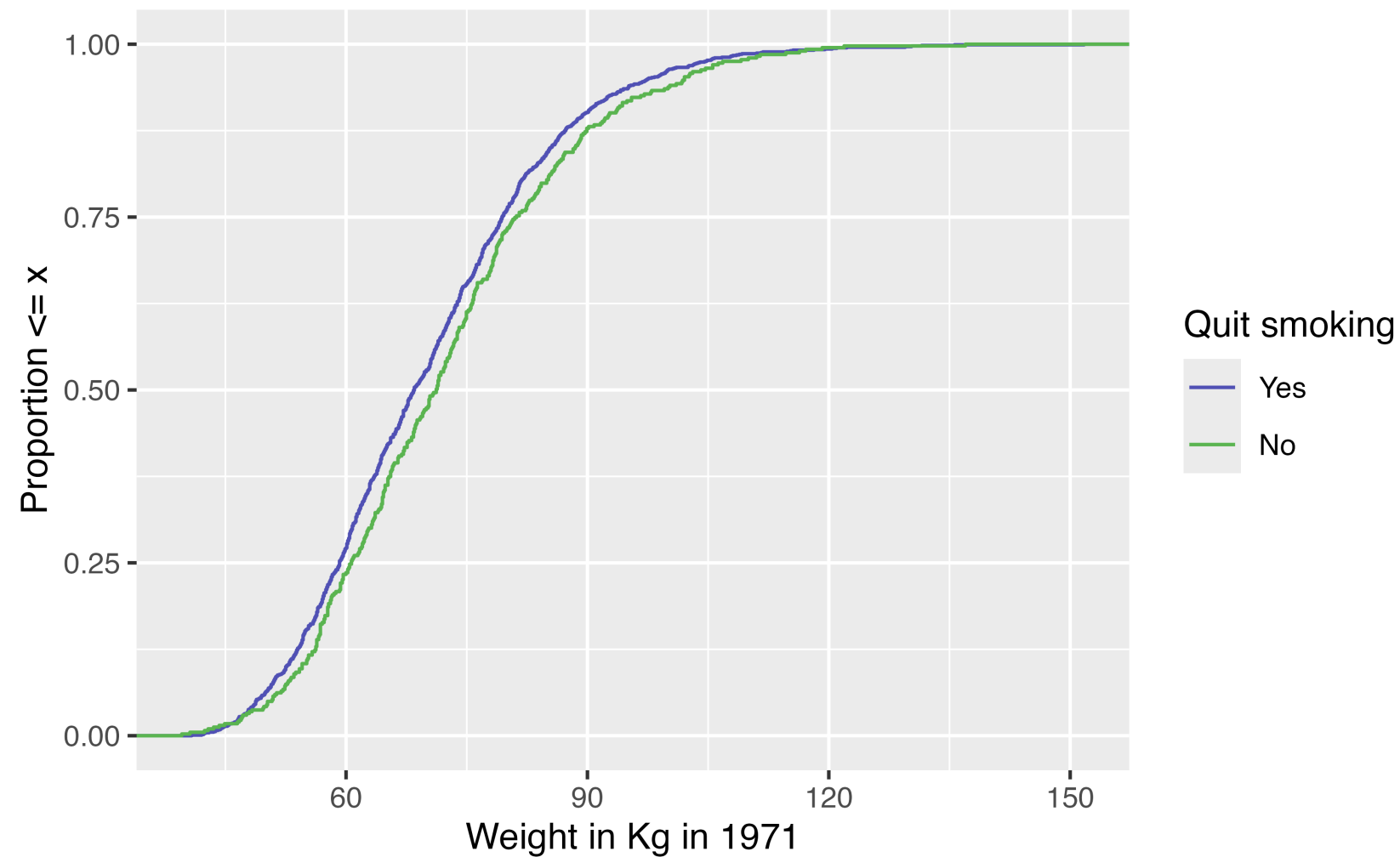
ECDF



Unweighted ECDF

```
1 ggplot(nhefs_complete_wts, aes(x = wt71, color = factor(qsmk))) +  
2   geom_ecdf() +  
3   scale_color_manual(  
4     "Quit smoking",  
5     values = c("#5154B8", "#5DB854"),  
6     labels = c("Yes", "No")  
7   ) +  
8   xlab("Weight in Kg in 1971") +  
9   ylab("Proportion <= x")
```

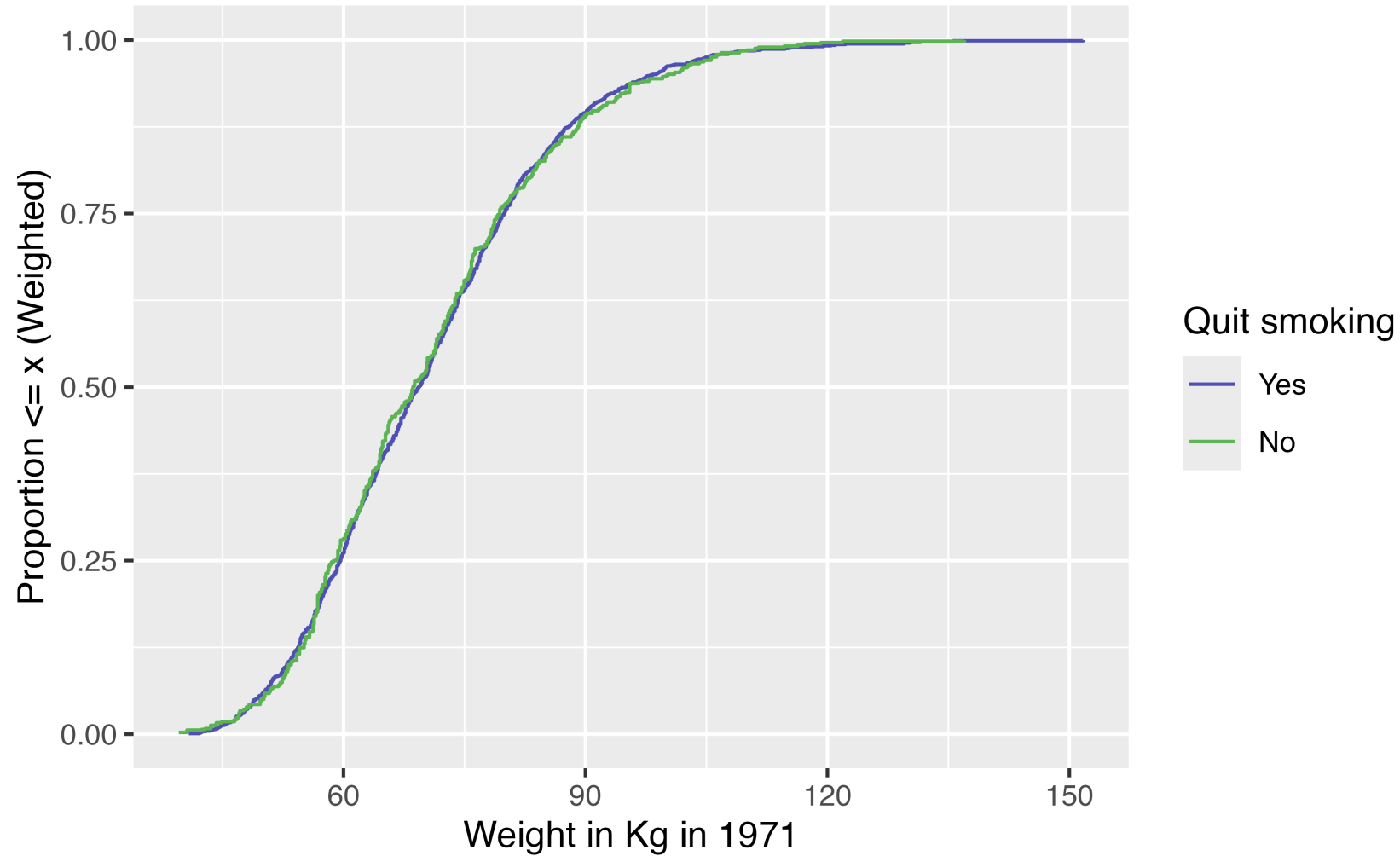
Unweighted ECDF



Weighted ECDF

```
1 ggplot(nhefs_complete_wts, aes(x = wt71, color = factor(qsmk))) +  
2   geom_ecdf(aes(weights = w_ate)) +  
3   scale_color_manual(  
4     "Quit smoking",  
5     values = c("#5154B8", "#5DB854"),  
6     labels = c("Yes", "No")  
7   ) +  
8   xlab("Weight in Kg in 1971") +  
9   ylab("Proportion <= x (Weighted)")
```

Weighted ECDF



Your turn 2

Create an unweighted ECDF examining the **park_temperature_high** confounder by whether or not the day had Extra Magic Hours.

Create a weighted ECDF examining the **park_temperature_high** confounder

***Bonus!* Weighted Tables in R**

1. Create a “design object” to incorporate the weights

```
1 library(survey)
2
3 svy_des <- svydesign(
4   ids = ~ 1,
5   data = df,
6   weights = ~ wts
7 )
```

2. Pass to

gtsummary::tbl_svysummary()

```
1 library(gtsummary)
2 tbl_svysummary(svy_des, by = x) |>
3   add_difference(everything() ~ "smd")
4 # modify_column_hide(ci) to hide CI column
```

Characteristic	0 N = 1,565 ¹	1 N = 1,561 ¹	Difference ²	95% CI ²
WEIGHT IN KILOGRAMS IN 1971	69 (60, 80)	69 (59, 79)	0.01	-0.06, 0.08
0: WHITE 1: BLACK OR OTHER IN 1971			0.01	-0.06, 0.08
0	1,359 (87%)	1,352 (87%)		
1	206 (13%)	209 (13%)		
AGE IN 1971	43 (33, 52)	43 (33, 53)	-0.01	-0.08, 0.06
0: MALE 1: FEMALE			0.00	-0.07, 0.07
0	764 (49%)	764 (49%)		
1	802 (51%)	797 (51%)		
NUMBER OF CIGARETTES SMOKED PER DAY IN 1971	20 (10, 25)	20 (10, 30)	0.02	-0.05, 0.09
YEARS OF SMOKING	24 (15, 33)	24 (14, 33)	0.00	-0.07, 0.07
IN RECREATION, HOW MUCH EXERCISE? IN 1971, 0:much exercise,1:moderate exercise,2:little or no exercise			0.04	-0.03, 0.11
0	302 (19%)	294 (19%)		
1	665 (42%)	691 (44%)		
2	599 (38%)	576 (37%)		
IN YOUR USUAL DAY, HOW ACTIVE ARE YOU? IN 1971, 0:very active, 1:moderately active, 2:inactive			0.03	-0.04, 0.10
0	700 (45%)	684 (44%)		
1	718 (46%)	738 (47%)		
2	147 (9.4%)	138 (8.9%)		

¹ Median (Q1, Q3); n (%)

² Standardized Mean Difference

Abbreviation: CI = Confidence Interval

Bonus Your Turn: **Weighted Tables**