Suplementary Material

Ranni Tewfik

2024-04-05

# 1. Documentation for Supplementary Material

The R codes and data files required to reproduce this analysis are available on [GitHub](https://github.com/rtewfik/Tewfik-MADA-project)

* The data folder contains the raw data (data/raw-data subfolder) and the processed data (data/processed-data subfolder).
* The code folder contains the R scripts for data processing (code/processing-code subfolder) and data analysis (code/analysis-code subfolder).
* The results folder contains code output, including tables (results/tables subfolder) and figures (results/figures subfolder).
* The products folder contains the project manuscript and the supplementary material.

The scripts in the GitHub repository should be run in the following order:

1. processing-code.qmd
2. exploratory-analysis.qmd
3. statistical-analysis.qmd

Load the following R packages before running any code: \* here \* dplyr \* skimr \* survival \* survminer \* ggplot2 \* tidyverse \* gtsummary \* knitr \* car

# 2. Distribution of Survival Time

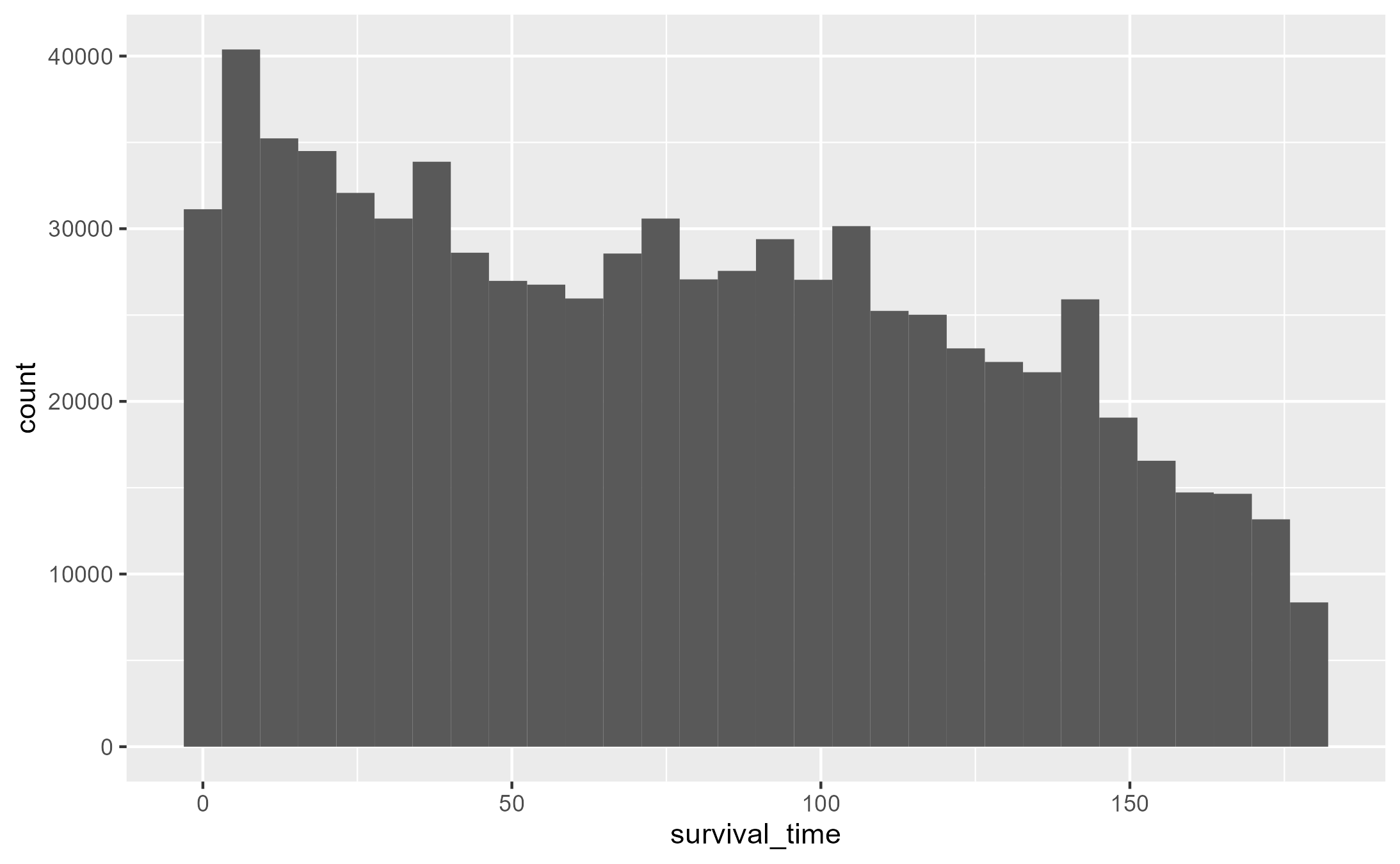
Figure 1. Histogram of survival time distribution in months.

library(here)

Warning: package 'here' was built under R version 4.3.2

here() starts at C:/Users/ranni/Desktop/MADA/Tewfik-MADA-project

library(knitr)  
knitr::include\_graphics(here("results/figures/survtime\_distribution.png"))

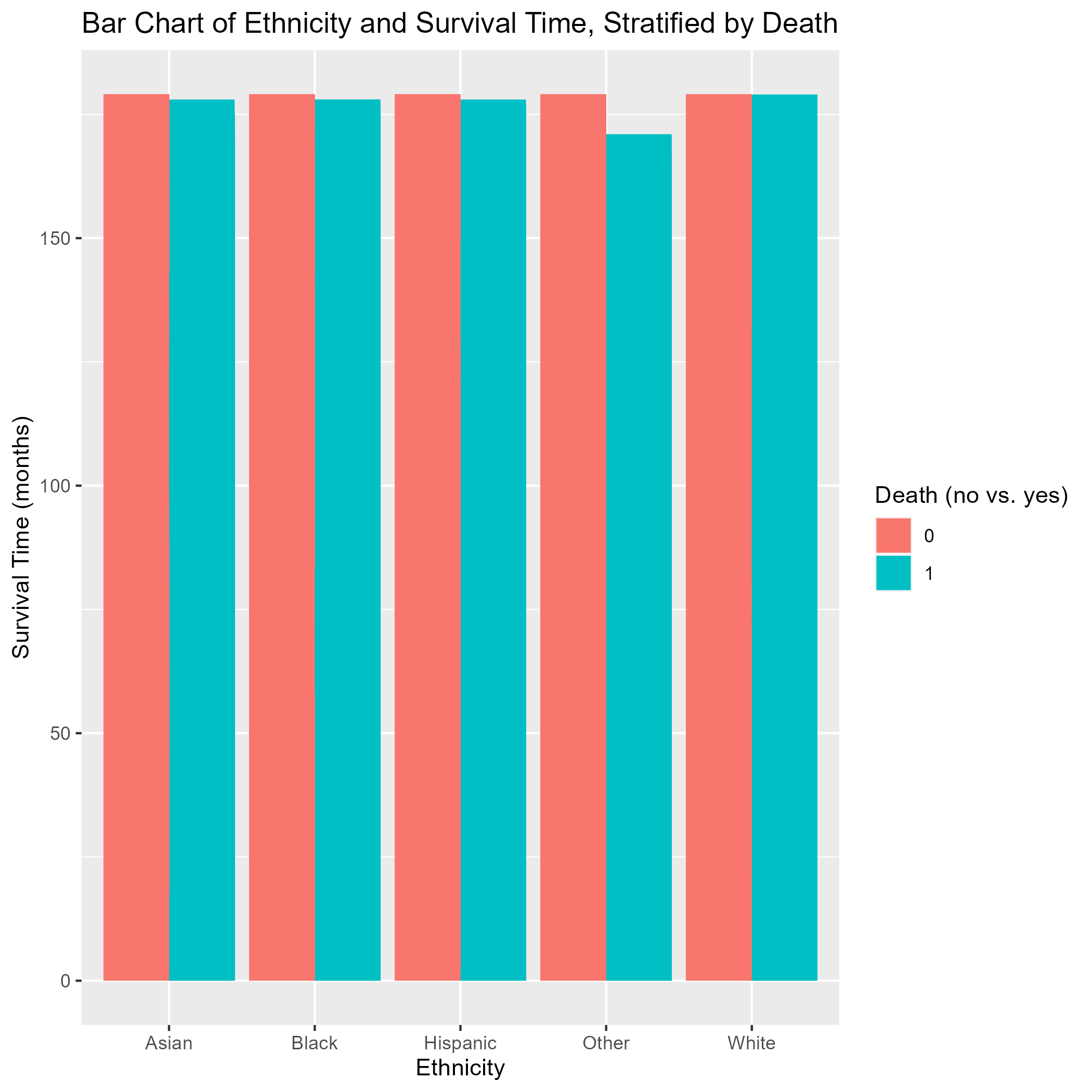


The histogram for survival time in months shows a right-skewed (positively skewed) distribution in which the mean is greater than the median.

# 3. Ethnicity and Survival Time, Stratified by Mortality

Figure 2. Bar Chart of Ethnicity and Survival Time, Stratified by Mortality

knitr::include\_graphics(here("results/figures/ethnicity\_survtime\_death\_barchart.png"))

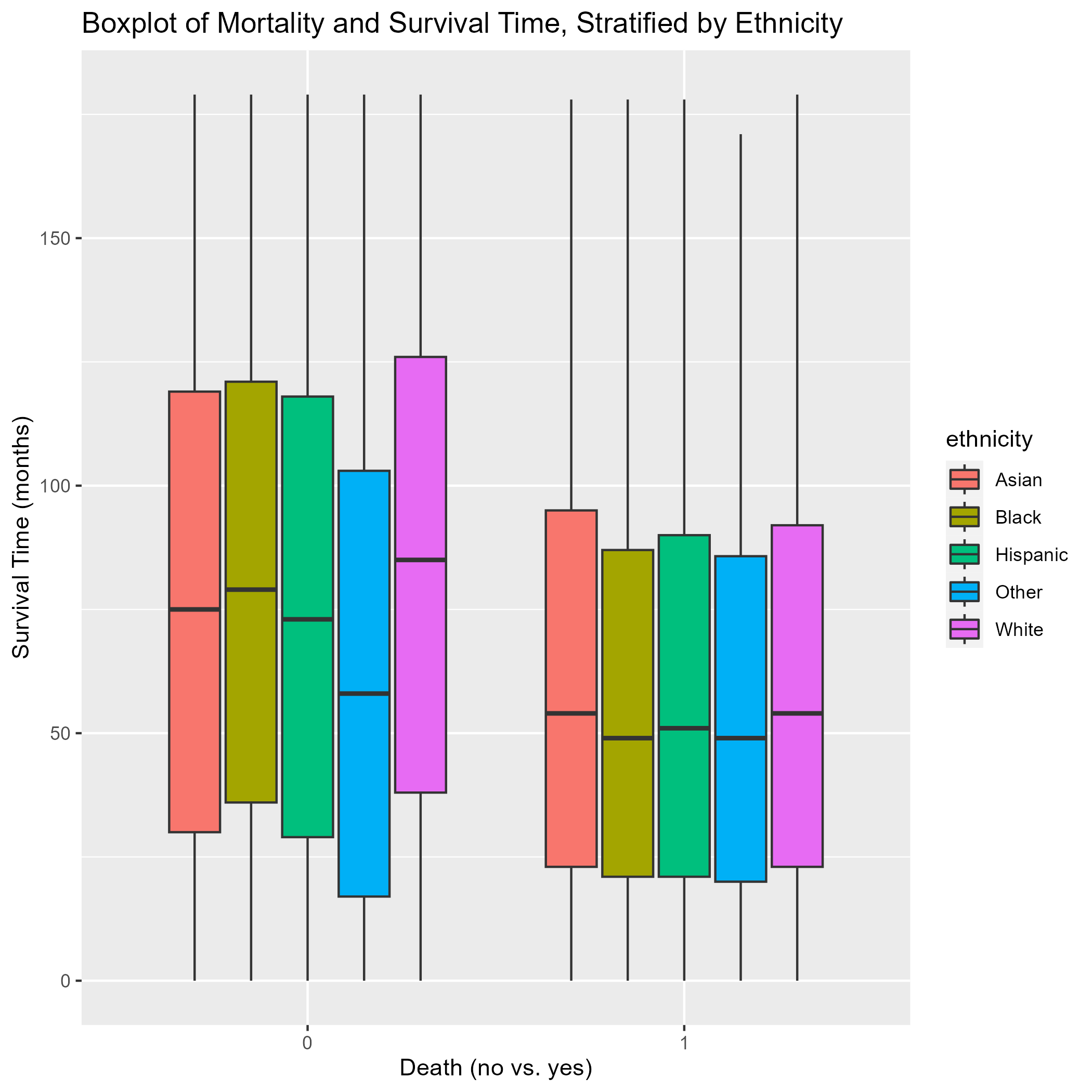


The bar chart shows that for each ethnicity, the survival time in months was slightly lower among those with all-cause mortality (vs. no death) during the follow-up period of the prostate cancer patients. The difference in survival time is largest among the “Other” group.

# 4. Mortality and Survival Time, Stratified by Ethnicity

Figure 3. Boxplot of Mortality and Survival Time, Stratified by Ethnicity.

knitr::include\_graphics(here("results/figures/death\_survtime\_ethnicity\_boxplot.png"))



The boxplot shows that among those with all-cause mortality, the survival time in months is lower across all ethnicities compared to those without all-cause mortality. The survival time is about the same in all ethnicities among those with all-cause mortality, whereas the survival time is noticeable lower among “Other” and higher among “White” among thise without all-cause mortality.

# 5. Survival Curves

Figure 4. Kaplan-Meier survival curves for males in the U.S. with prostate cancer, stratified by age group.

knitr::include\_graphics(here("results/figures/survivalcurve\_age.png"))

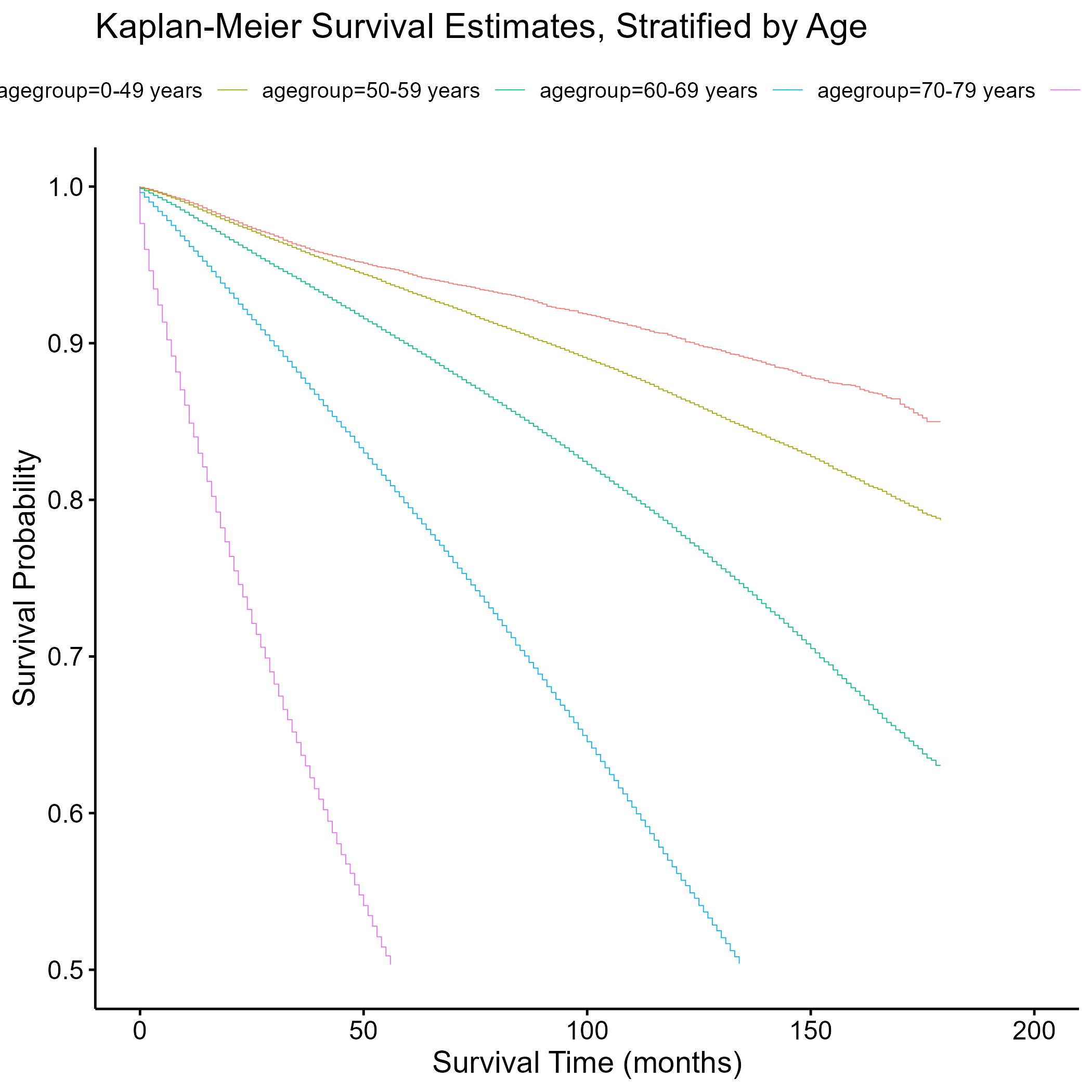


Figure 5. Kaplan-Meier survival curves for males in the U.S. with prostate cancer, stratified by cancer stage.

knitr::include\_graphics(here("results/figures/survivalcurve\_stage.png"))

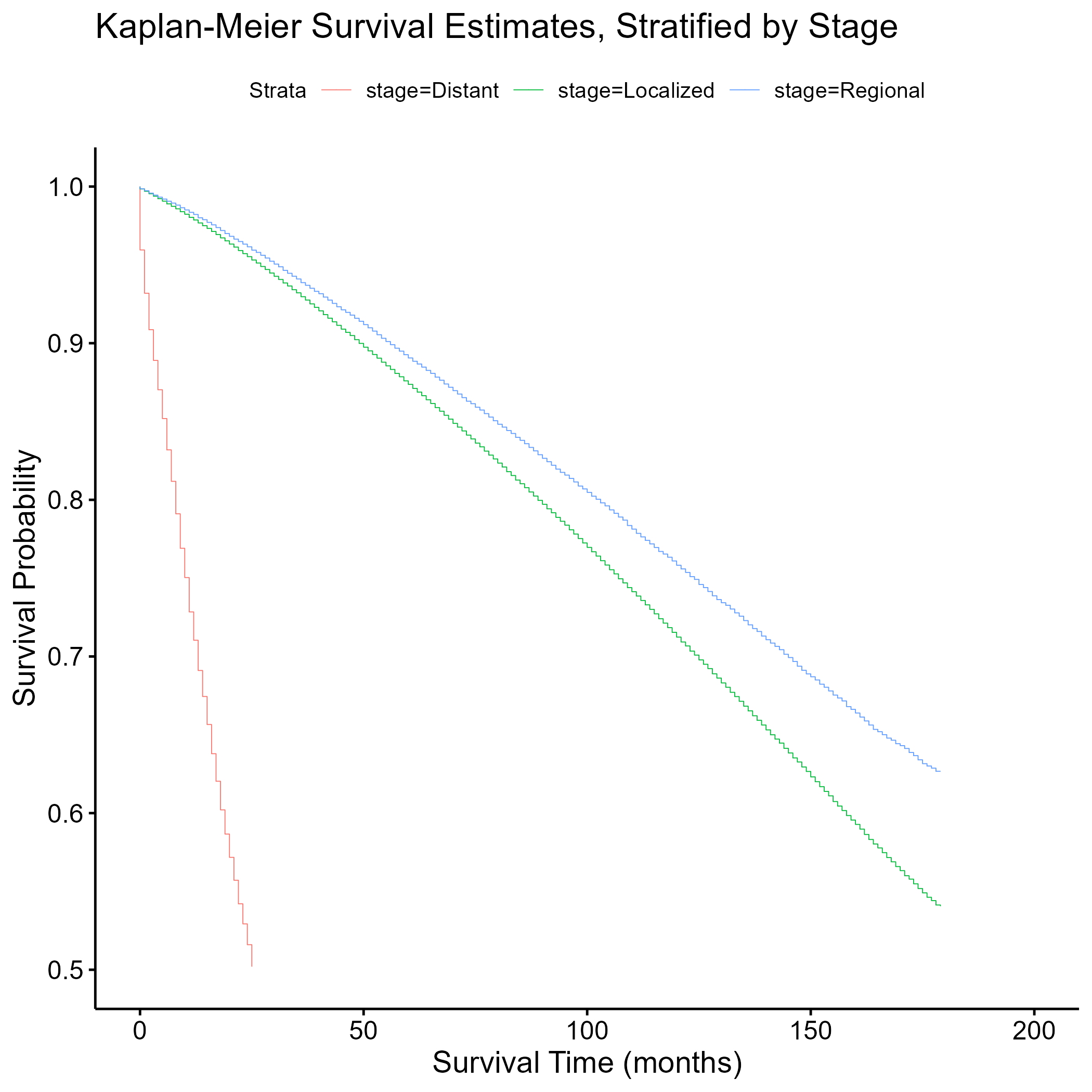


Figure 6. Kaplan-Meier survival curves for males in the U.S. with prostate cancer, stratified by education.

knitr::include\_graphics(here("results/figures/survivalcurve\_education.png"))

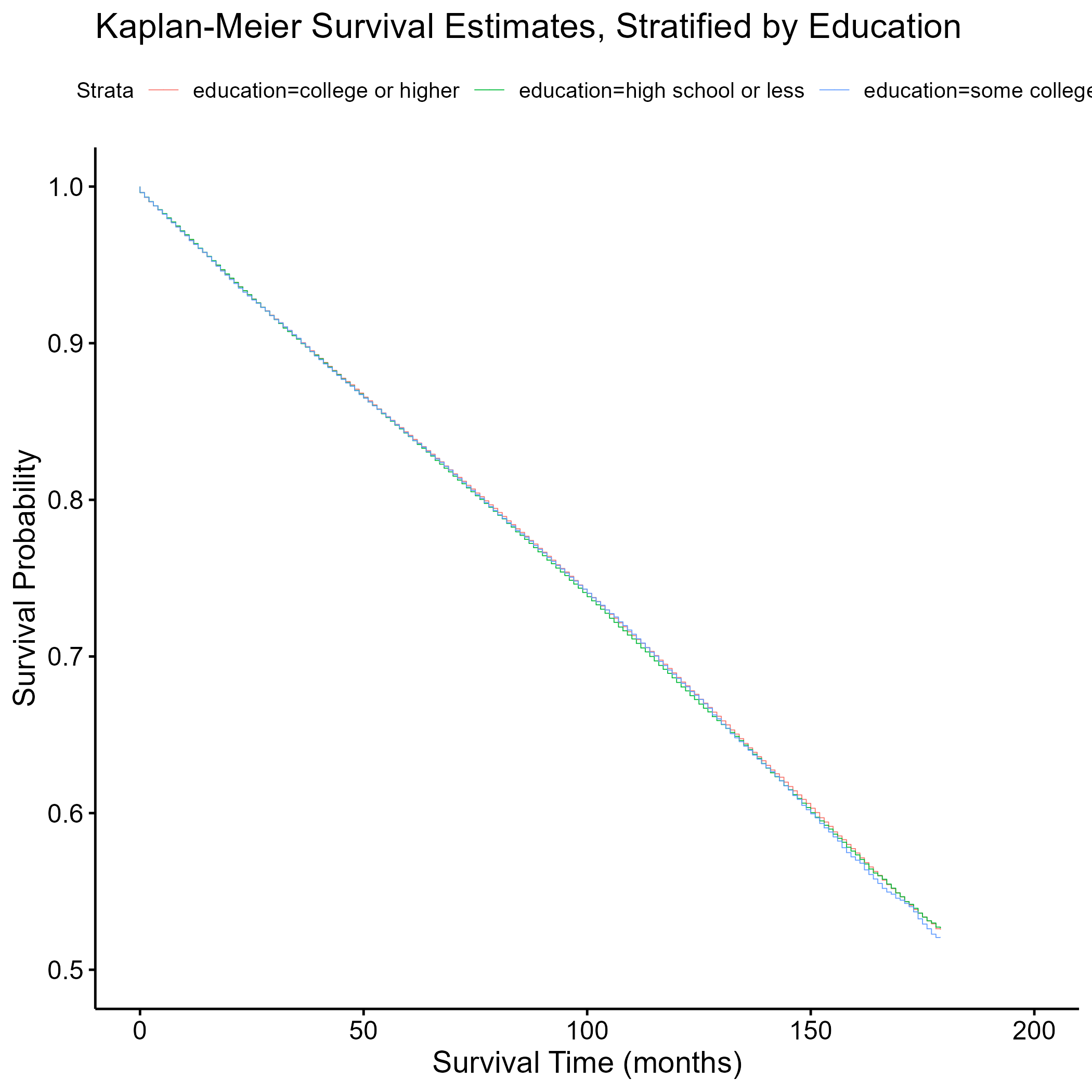
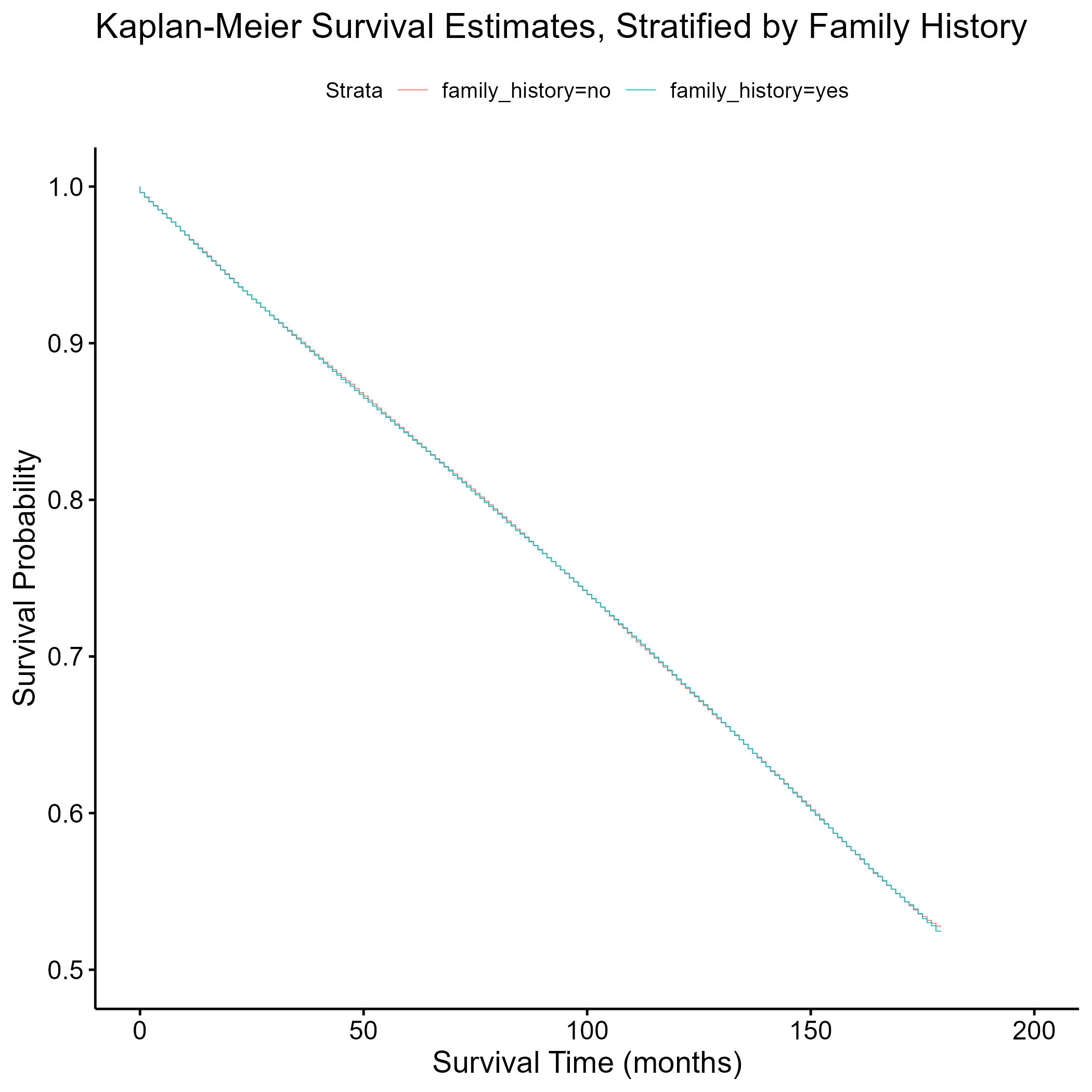


Figure 7. Kaplan-Meier survival curves for males in the U.S. with prostate cancer, stratified by family history of cancer.

knitr::include\_graphics(here("results/figures/survivalcurve\_fhx.png"))



The survival curves show that older prostate cancer patients with distant cancer stage had lower survival than their younger counterparts with localized or regional stage. There seemed to be no difference in survival among the different educational attainment groups or family history of cancer groups.

# 6. Cox Proportional Hazards Assessment

Figure 8. Log-log survival curve for ethnicity.

knitr::include\_graphics(here("results/figures/loglogcurve\_ethnicity.png"))

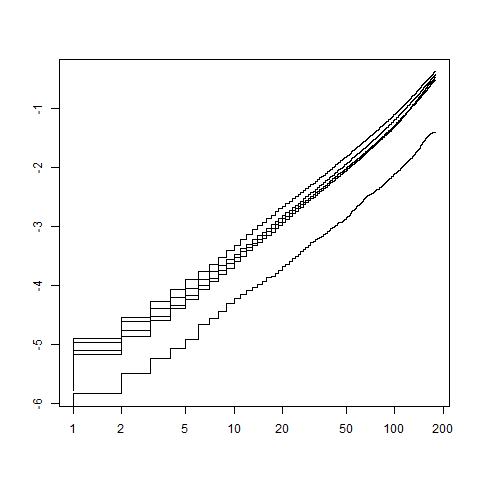


Figure 9. Log-log survival curve for age group.

knitr::include\_graphics(here("results/figures/loglogcurve\_age.png"))

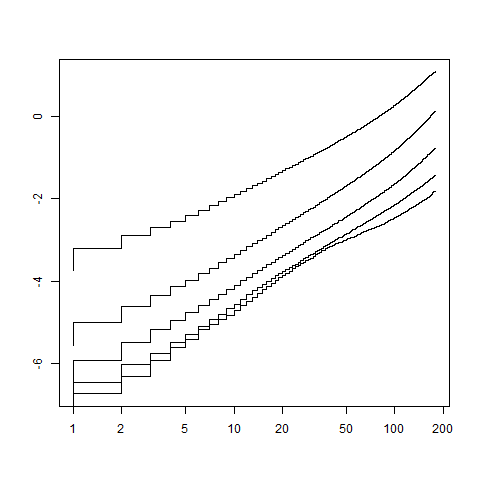
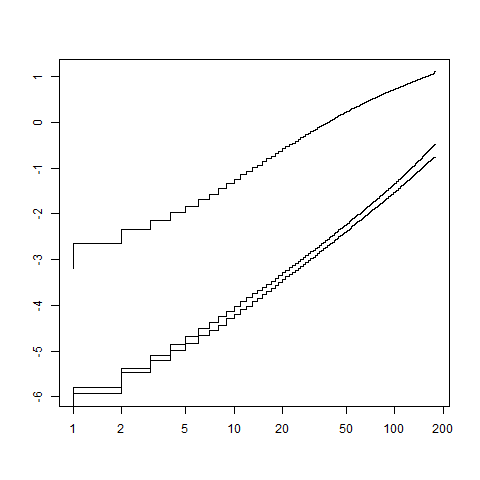


Figure 10. Log-log survival curve for cancer stage.

knitr::include\_graphics(here("results/figures/loglogcurve\_stage.png"))



The log-log survival curves are parallel for “ethnicity”, “agegroup”, and “stage”, indicating that the Cox proportional hazards assumption is met for those variables.

# 7. Unadjusted Cox Proportional Hazards Model

Table 1. Unadjusted hazard ratios and 95% confidence intervals for the effect of ethnicity on time to all-cause mortality among males in the U.S. with prostate cancer (SEER database).

|  | Characteristic | Hazard.ratio | X95..CI..lower. | X95..CI..upper. |
| --- | --- | --- | --- | --- |
| relevel(factor(ethnicity), ref = “White”)Asian | Asian | 0.9164256 | 0.8959684 | 0.9373499 |
| relevel(factor(ethnicity), ref = “White”)Black | Black | 1.0958792 | 1.0820464 | 1.1098889 |
| relevel(factor(ethnicity), ref = “White”)Hispanic | Hispanic | 0.9177404 | 0.9022775 | 0.9334682 |
| relevel(factor(ethnicity), ref = “White”)Other | Other | 0.4029607 | 0.3816502 | 0.4254612 |

The rate of all-cause mortality was 1.10 times higher in Non-Hispanic Black males than Non-Hispanic White males (HR = 1.10; 95% CI = 1.08 – 1.11). Compared to Non-Hispanic White males, the hazard ratios were less than 1.00 and statistically significant for males who were Non-Hispanic Asian/Pacific Islander, Hispanic, and other race/ethnicity.

Table 2. Unadjusted hazard ratios and 95% confidence intervals for the effect of age on time to all-cause mortality among males in the U.S. with prostate cancer (SEER database).

|  | Characteristic | Hazard.ratio | X95..CI..lower. | X95..CI..upper. |
| --- | --- | --- | --- | --- |
| relevel(factor(agegroup), ref = “0-49 years”)50-59 years | 50-59 years | 1.396145 | 1.328893 | 1.466801 |
| relevel(factor(agegroup), ref = “0-49 years”)60-69 years | 60-69 years | 2.424444 | 2.311338 | 2.543085 |
| relevel(factor(agegroup), ref = “0-49 years”)70-79 years | 70-79 years | 5.506989 | 5.251173 | 5.775267 |
| relevel(factor(agegroup), ref = “0-49 years”)80+ years | 80+ years | 16.633487 | 15.855086 | 17.450103 |

The rate of all-cause mortality was 16.63 times higher in males age 80+ years than males younger than 50 years (HR = 16.63; 95% CI = 15.86 – 17.45). Compared to males younger than 50 years, the hazard ratios were greater than 1.00 and statistically significant for males 50 years or older.

Table 3. Unadjusted hazard ratios and 95% confidence intervals for the effect of cancer stage on time to all-cause mortality among males in the U.S. with prostate cancer (SEER database).

|  | Characteristic | Hazard.ratio | X95..CI..lower. | X95..CI..upper. |
| --- | --- | --- | --- | --- |
| relevel(factor(stage), ref = “Localized”)Distant | Distant | 10.0571418 | 9.9301736 | 10.1857336 |
| relevel(factor(stage), ref = “Localized”)Regional | Regional | 0.8141635 | 0.8013803 | 0.8271507 |

The rate of all-cause mortality was 10.06 times higher in males with distant cancer stage than males with localized cancer stage (HR = 10.06; 95% CI = 9.93 - 10.19). Compared to males with localized cancer stage, the rate of all-cause mortality was 0.81 times the corresponding rate in males with regional cancer stage (HR = 0.81; 95% CI = 0.80 - 0.83).