PSTAT 175 Project: Term Lengths of Canadian Senators in Office

Hannah Mok, Caleb Mazariegos, Joshua Price

2022-12-11

I. Introduction

For this survival analysis project, we will analyze the lengths of Canadian senators' terms in office. The data for this project is sourced from the Python package "lifelines: Survival Analysis in Python" by Davidson-Pilon (2019). This dataset contains ten variables: name of the senator, political affiliation at appointment, province/territory, the person who advised the senator's appointment, their term in office (yyyy.mm.dd), start date, end date, reason for ending term, the difference between start date and end date in days, and an indicator variable to show whether the end of the senator's term was observed or not. Six of the variables are factors that describe the situation around the senator's appointment and term in office, and the other four are date objects describing the dates and number of days in office.

The main goal of our project is to identify how long Canadian senators are able to stay in office before ending their term due to death, retirement, resignation, or declining the appointment. To do this, we aim to build a best-fit model by selecting the most significant covariates, thus observing which of the covariates have the largest effect on Canadian senators' term lengths. We will choose out of four possible covariates, which are all factors: political affiliation at appointment, province or territory, who they were appointed on the advice of, and reason for ending their term. The political affiliation at appointment variable refers to the political party that the senator was affiliated with when being appointed. Although Canada currently only has five federal parties, this variable includes 17 names of parties that have risen and declined over the years. We choose to only analyze the three major parties in Canadian history: Conservative (1867-1942), Liberal Party of Canada, and the Progressive Conservative Party. Next, the province or territory variable describes the province or territory of Canada that the senator is representing. The third possible variable, who the senator was appointed on the advice of, refers to the person or entity that advised the appointment of the senator. On top of appointments advised by the Royal Proclamation, there are 21 individuals who advised appointments. The last variable is the reason for the senator ending their term in office. This includes: death, retirement, resignation, or declining the appointment. The failure time variable we are using is `diff_days`. This variable is the total number of days the senator was in office before ending their term.

By building a survival analysis, we can identify the most relevant variables to the number of days that a senator is able to stay in office. Then, we can make conclusions on what may influence or affect a senator to leave office, and provide a direction for future research into the covariates or reasons that significantly affect the length of time that a Canadian senator serves in office.

In our project, we use the `tidyverse` library to handle data and `ggplot` package for plotting.

Load Data

We first load the Canadian senator term data from the lifelines library in Python. This dataset was pulled from the library and downloaded as a csv file using Python. The documentation can be found on "lifelines: Survival Analysis in Python" (2019).

```r
senators.raw <- read_csv("lifelines_canadian_senators.csv")
```

```
## Rows: 933 Columns: 10
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (6): Name, Political Affiliation at Appointment, Province / Territory, ...
## dbl  (1): diff_days
## lgl  (1): observed
## dttm (2): start_date, end_date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
senators.raw <- clean_names(senators.raw)
summary(senators.raw)
```

```
##      name           political_affiliation_at_appointment province_territory
##  Length:933         Length:933                           Length:933
##  Class :character   Class :character                     Class :character
##  Mode  :character   Mode  :character                     Mode  :character
##
##
##
##  appointed_on_the_advice_of term_yyyy_mm_dd      start_date
##  Length:933                 Length:933         Min.   :1867-10-23 00:00:00
##  Class :character           Class :character   1st Qu.:1901-04-04 00:00:00
##  Mode  :character           Mode  :character   Median :1942-11-19 00:00:00
##                                                Mean   :1940-06-22 16:35:30
##                                                3rd Qu.:1979-10-03 00:00:00
##                                                Max.   :2013-03-25 00:00:00
##     end_date                        reason            diff_days
##  Min.   :1867-10-23 00:00:00   Length:933         Min.   :    0
##  1st Qu.:1916-08-25 00:00:00   Class :character   1st Qu.: 2082
##  Median :1957-03-08 00:00:00   Mode  :character   Median : 4578
##  Mean   :1954-04-06 15:50:58                      Mean   : 5036
##  3rd Qu.:1997-06-20 00:00:00                      3rd Qu.: 7331
##  Max.   :2013-10-01 22:20:10                      Max.   :17731
##   observed
##  Mode :logical
##  FALSE:99
##  TRUE :834
##
##
##
```

Data Cleaning

We ensure that there are no NA values in our possible covariates except `reason`.

```r
senators <- senators.raw %>% drop_na(name,
                                     political_affiliation_at_appointment,
```

```
                                        province_territory,
                                        appointed_on_the_advice_of,
                                        term_yyyy_mm_dd, start_date, end_date,
                                        diff_days, observed)
```

We then check the reasons for ending the terms:

```
unique(senators$reason)
```

```
## [1] "Death"                  "Retirement"            "Resignation"
## [4] NA                       "Appointment declined"
```

There are 99 senators who do not have an available reason for leaving the office. We will do a standard Cox model analysis on the original dataset first, and then drop the observations with no available reason for the competing risk analysis.

We also drop the 3 observations with a reason of "Appointment declined", since they were not in office and their recorded begin date and end date are the same day.

```
senators <- senators %>% filter(reason!="Appointment declined"|is.na(reason))
```

After looking through the data, we see that terms that do not have an observed end date yet (i.e. observed=FALSE) have an end_date of "2013-10-01 22:20:10 UTC" and do not have a reason for ending the term (i.e. reason=NA). There are 99 terms that do not have an observed end date. We add a new column, status, to recode the observed values as 0 for FALSE and 1 for TRUE. We also recode the NA reason to "Unobserved".

```
# Add status column to recode observed values as 0 and 1 otherwise
senators = senators %>%
  mutate(status = ifelse(observed=="TRUE", 1, 0))
# Recode NA reasons to Unobserved
senators$reason[is.na(senators$reason)] = "Unobserved"
```

We also add a column for the year of the start date to determine if there was a time or historical effect on survival.

```
senators = senators %>%
  mutate(start_year = lubridate::year(start_date))
```

Then, we look at the frequencies of each political affiliation.

```
affcount <- senators %>%
  group_by(political_affiliation_at_appointment) %>%
  summarise(counts = n())
affcount
```

```
## # A tibble: 17 x 2
##    political_affiliation_at_appointment counts
##    <chr>                                 <int>
##  1 Conservative (1867-1942)                210
##  2 Conservative Party of Canada             61
```

```
##   3 Independent                          11
##   4 Independent Conservative              4
##   5 Independent Liberal                   5
##   6 Liberal Party of Canada             478
##   7 Liberal Progressive                   1
##   8 Liberal-Conservative                 37
##   9 Nationalist                           2
##  10 Nationalist Conservative              1
##  11 Nationalist Liberal                   2
##  12 New Democratic Party                  1
##  13 Progressive Conservative Party      112
##  14 Reform Party                          1
##  15 Reformer                              1
##  16 Social Credit Party                   1
##  17 Unionist (Liberal)                    2
```

To give us a more meaningful analysis that focuses on the most ubiquitous ideologies or affiliations in Canada's political history, we looked into the historical context of Canada's political parties.

The Conservative Party of Canada (1867-1942) became the Unionist Party from 1917 to 1920, and then was later renamed the Liberal-Conservative Party from 1921 to 1938. The Progressive Conservative Party is the revised name for the original Conservative Party of Canada that was implemented in 1942. In 2003, this party combined with the Reform Party (later known as Canadian Alliance) to form the current Conservative Party of Canada. As such, these parties were all recoded to be part of the "Conservative" category.

The Liberal Party of Canada originated from the mid-19th century Reformers. As such, they were included as "Liberal".

All other affiliations only had a few observations and are made up of smaller parties, so they were recoded as "Other".

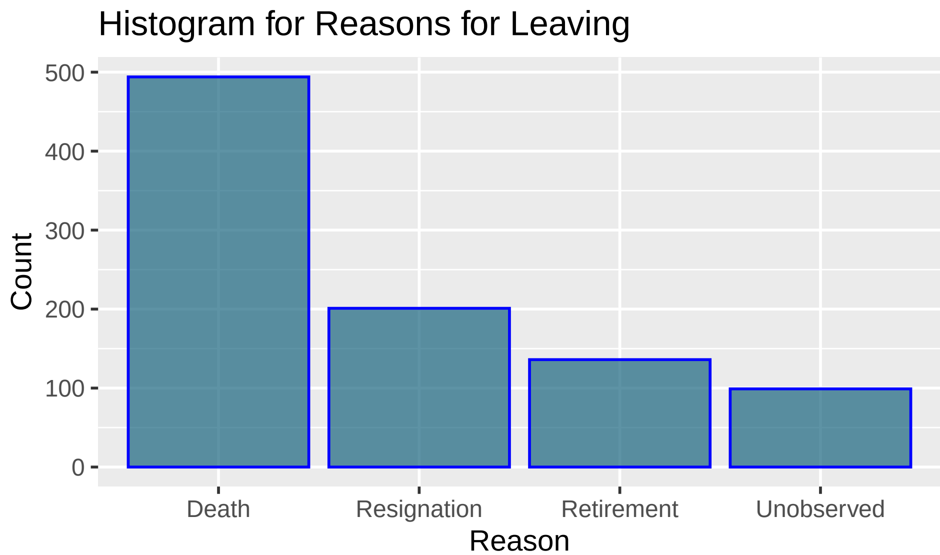
```
senators <- senators %>%
  mutate(affiliation =
       case_when(political_affiliation_at_appointment == "Conservative (1867-1942)"|
       political_affiliation_at_appointment == "Unionist (Liberal)"|
       political_affiliation_at_appointment == "Liberal-Conservative"|
       political_affiliation_at_appointment == "Progressive Conservative Party"|
       political_affiliation_at_appointment == "Reform Party"|
       political_affiliation_at_appointment == "Conservative Party of Canada"
       ~ "Conservative",
       political_affiliation_at_appointment == "Liberal Party of Canada"|
       political_affiliation_at_appointment == "Reformer"
       ~ "Liberal",
       TRUE ~ "Other"))
```

Graphs

Before we dive into analysis, we take a preliminary look into the different buckets that each senator may classify as. We first build histograms for our different covariates to derive some insights into their distributions.

Reasons

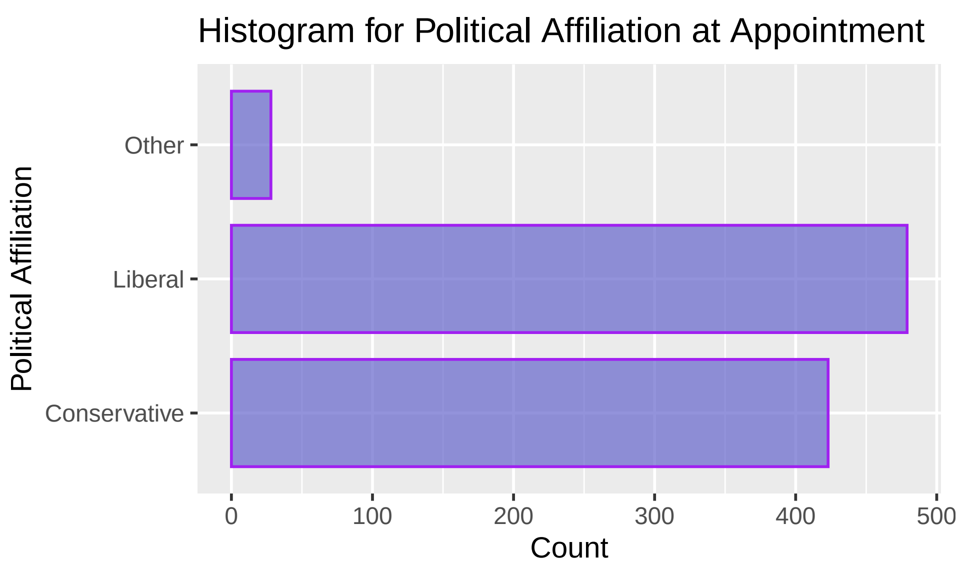
First, we look at the frequencies of each possible reason for leaving the office. Note that we include "Unobserved", though it represents the observations for which the study ended before the end of term was observed.



For this dataset, we see that death is one of the most observed reasons for leaving office. It is more than twice of the observed resignations and close to three times of the observed retirements.

Political Affiliation

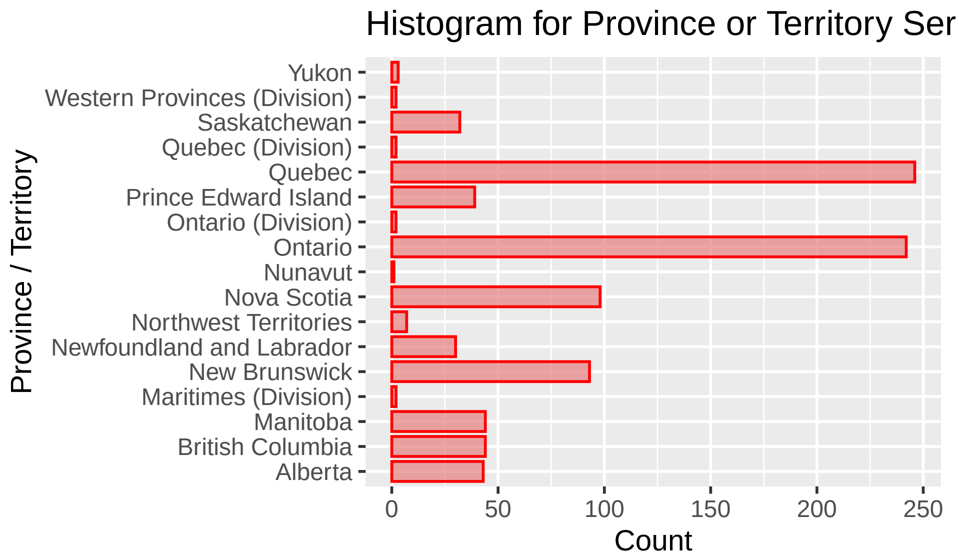
Next, we observe the number of senators within each political affiliation at appointment.



We observe that Liberal Party of Canada has the most observations, drastically more than Conservative (1867 to 1942) and Progressive Conservative.

Province / Territory

Then, we look at the distribution of provinces or territories served.



We note that Ontario and Quebec are highly represented, which is reasonable since they are both large and highly populated territories. Since we also observe several categories with minimal counts such as the Northwest, Nunavet, and Yukon Territories or the self-designated divisions of Maritimes, Ontario, Quebec, and the Western Provinces, we group these categories to avoid errors in our Cox models. We also group the smaller Provinces with fewer than 50 senators into the same category as the model still runs into errors when only the minimal counts are grouped. The new category is coded as "Other".

```
provcount <- senators %>%
  group_by(province_territory) %>%
  summarise(counts = n())
provcount
```

```
## # A tibble: 17 x 2
##    province_territory           counts
##    <chr>                         <int>
##  1 Alberta                          43
##  2 British Columbia                 44
##  3 Manitoba                         44
##  4 Maritimes (Division)              2
##  5 New Brunswick                    93
##  6 Newfoundland and Labrador        30
##  7 Northwest Territories             7
##  8 Nova Scotia                      98
##  9 Nunavut                           1
## 10 Ontario                         242
## 11 Ontario (Division)                2
## 12 Prince Edward Island             39
## 13 Quebec                          246
## 14 Quebec (Division)                 2
## 15 Saskatchewan                     32
## 16 Western Provinces (Division)      2
## 17 Yukon                             3
```

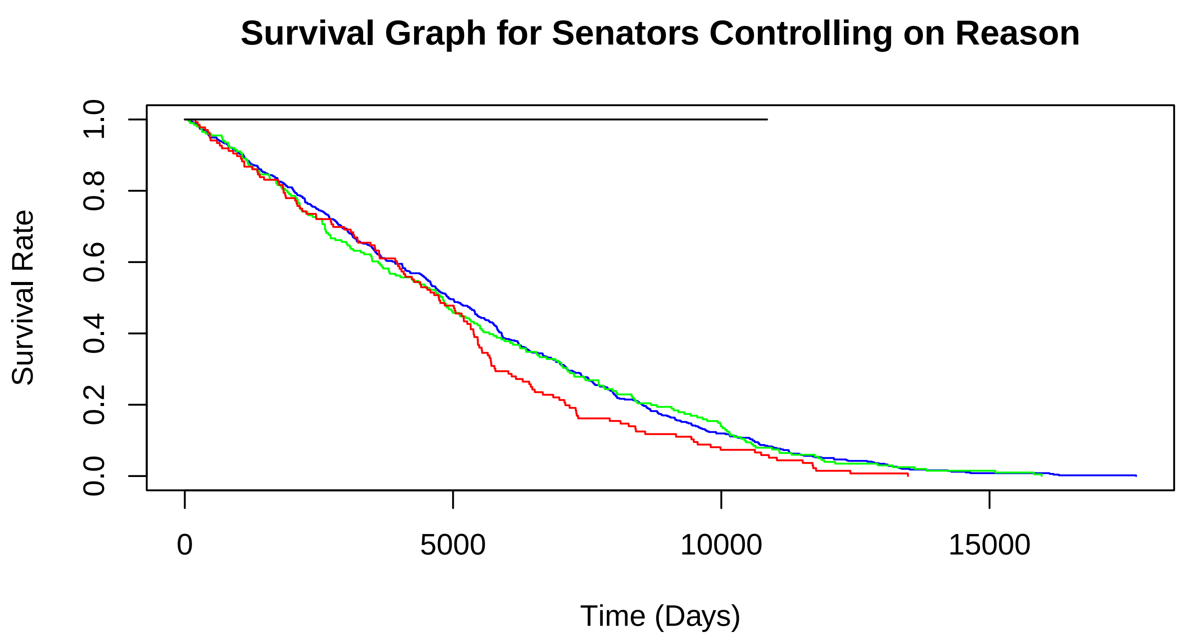
```
senators = senators %>%
  mutate(province_territory =
         case_when(province_territory == "Yukon"|
                   province_territory == "Nunavut"|
                   province_territory == "Western Provinces (Division)"|
                   province_territory == "Quebec (Division)"|
                   province_territory == "Ontario (Division)"|
                   province_territory == "Northwest Territories"|
                   province_territory == "Maritimes (Division)" |
                   province_territory == "Saskatchewan"|
                   province_territory == "Prince Edward Island"|
                   province_territory == "Newfoundland and Labrador"|
                   province_territory == "Manitoba"|
                   province_territory == "British Columbia"|
                   province_territory == "Alberta"
                   ~ "Other",
                   TRUE ~ province_territory))
```

We are left with 5 large categories: New Brunswick, Nova Scotia, Ontario, Quebec, and Other.

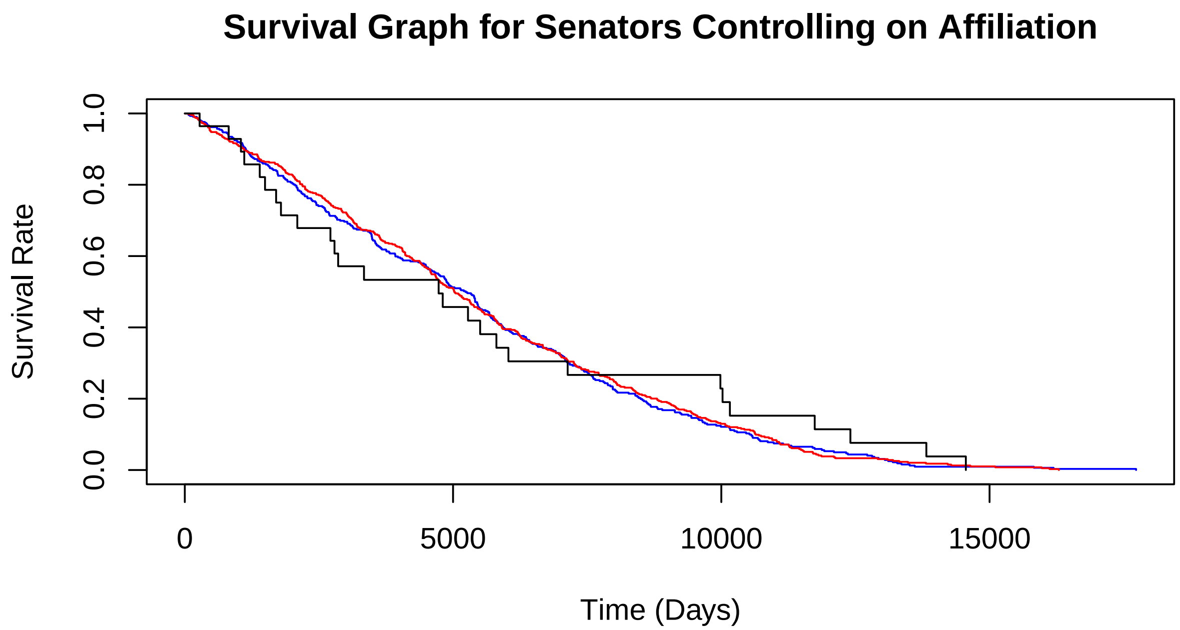
After looking into the frequencies of these covariates, we want to check the relationship between some of the covariates and the number of days in office.

Reason vs Number of days in office

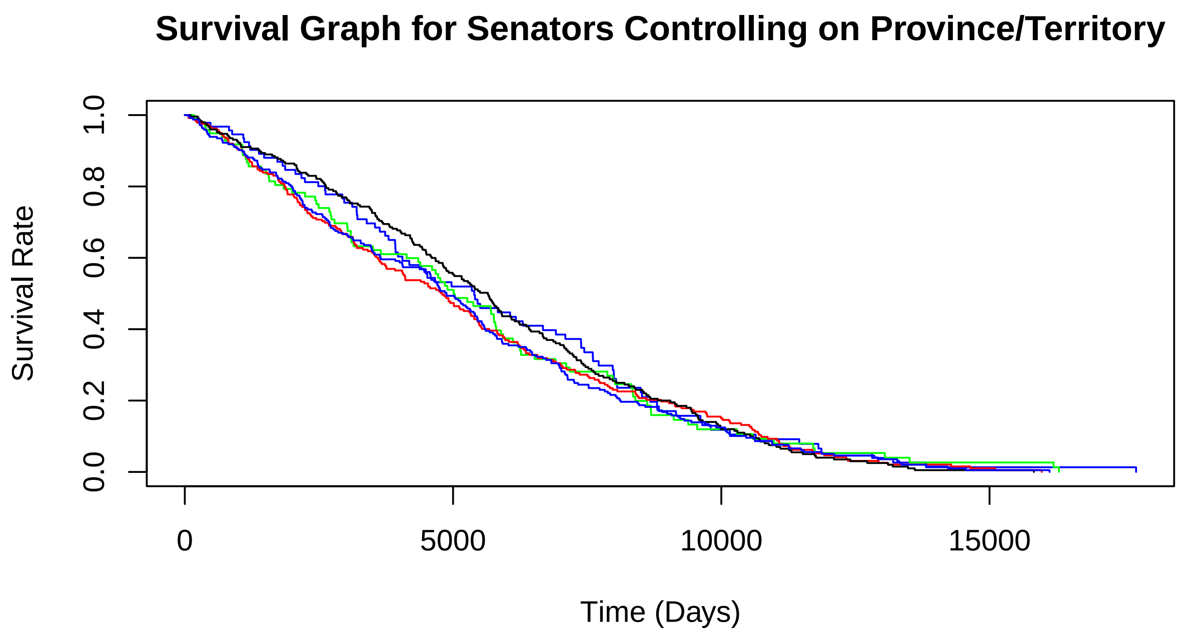
First, we look at the relationship between reason for leaving office and the number of days held in office.



Political Affiliation vs Number of days in office



Province / Territory vs Number of days in office



II. Model Fitting

Naive Standard Model using Cox

We use the Cox proportional hazards model because our data includes categorical values.
We fit a naive Cox model to our data:

```
surv_vec <- Surv(senators$diff_days, senators$status)
res.cox <- coxph(surv_vec ~ factor(affiliation) + factor(province_territory) +
                   start_year, data = senators)
summary(res.cox)
```

```
## Call:
## coxph(formula = surv_vec ~ factor(affiliation) + factor(province_territory) +
##     start_year, data = senators)
##
##   n= 930, number of events= 831
##
##                                              coef  exp(coef)   se(coef)      z
## factor(affiliation)Liberal             -0.0465870  0.9544815  0.0736137 -0.633
## factor(affiliation)Other               -0.1273141  0.8804570  0.2024945 -0.629
## factor(province_territory)Nova Scotia   0.0591509  1.0609353  0.1544940  0.383
## factor(province_territory)Ontario       0.1132099  1.1198670  0.1299788  0.871
## factor(province_territory)Other         0.0122689  1.0123444  0.1317441  0.093
## factor(province_territory)Quebec        0.1484515  1.1600365  0.1295529  1.146
## start_year                              0.0014431  1.0014441  0.0009033  1.598
##                                        Pr(>|z|)
## factor(affiliation)Liberal                0.527
## factor(affiliation)Other                  0.530
## factor(province_territory)Nova Scotia     0.702
## factor(province_territory)Ontario         0.384
## factor(province_territory)Other           0.926
## factor(province_territory)Quebec          0.252
## start_year                                0.110
##
##                                        exp(coef) exp(-coef) lower .95 upper .95
## factor(affiliation)Liberal                0.9545     1.0477    0.8262     1.103
## factor(affiliation)Other                  0.8805     1.1358    0.5920     1.309
## factor(province_territory)Nova Scotia     1.0609     0.9426    0.7838     1.436
## factor(province_territory)Ontario         1.1199     0.8930    0.8680     1.445
## factor(province_territory)Other           1.0123     0.9878    0.7820     1.311
## factor(province_territory)Quebec          1.1600     0.8620    0.8999     1.495
## start_year                                1.0014     0.9986    0.9997     1.003
##
## Concordance= 0.517  (se = 0.012 )
## Likelihood ratio test= 5.1  on 7 df,   p=0.6
## Wald test            = 5.08  on 7 df,   p=0.6
## Score (logrank) test = 5.09  on 7 df,   p=0.6
```

We observe that none of the covariates do not have significant p-values under .05 significance.

Stepwise Selection and AIC

Stepwise selection is a combination of both forward and backward selection. We start with no predictors, then sequentially add the most contributive predictors (like forward selection). After adding each new variable, remove any variables that no longer provide an improvement in the model fit (like backward selection). We choose the model with the lowest AIC.

```
# full model with possible covariates
fit1 <- coxph(surv_vec ~ start_year + factor(province_territory) +
                   factor(affiliation), data = senators)
# intercept only model
fit2 <- coxph(surv_vec ~ 1, senators)

# perform stepwise regression
stepwise = stepAIC(fit2, direction = "both",
                   scope = list(upper=fit1, lower=fit2))
```

```
## Start:  AIC=9611.91
## surv_vec ~ 1
##
##                              Df    AIC
## <none>                          9611.9
## + start_year                  1 9612.2
## + factor(affiliation)         2 9615.6
## + factor(province_territory)  4 9617.7
```

Our model indicates that the null model has the lowest AIC and thus is the best model. However, for the remainder of our analyses we will still use start_year to illustrate process of survival analysis. Its covariate was also the lowest and closest to being significant under a 0.10 significance level, so it may still be meaningful to consider it in our model.

Fitted Model using Cox

We use the Cox proportional hazards model because our data includes categorical values.

```
res.cox2 <- coxph(Surv(diff_days, status) ~ start_year, data = senators)
summary(res.cox2)
```

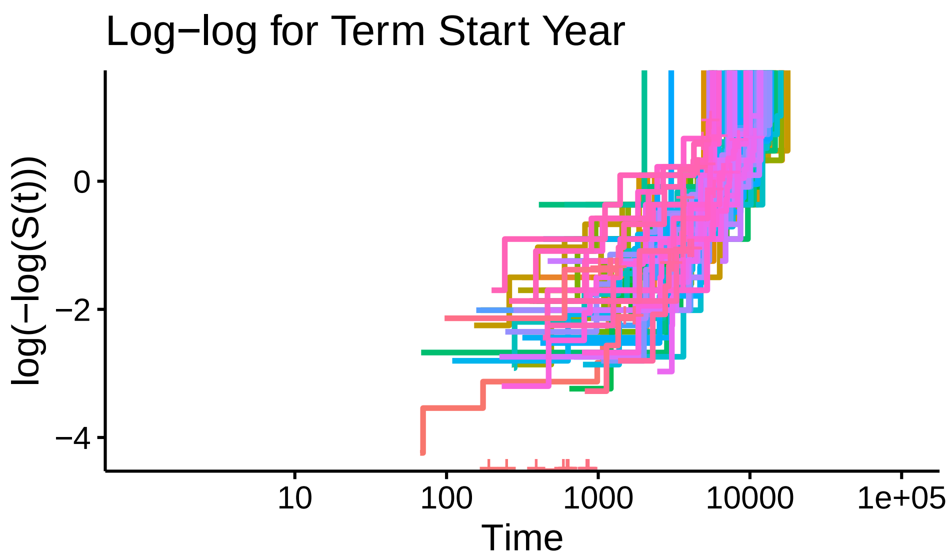
```
## Call:
## coxph(formula = Surv(diff_days, status) ~ start_year, data = senators)
##
##   n= 930, number of events= 831
##
##                 coef exp(coef)  se(coef)     z Pr(>|z|)
## start_year 0.0011327 1.0011334 0.0008611 1.315    0.188
##
##            exp(coef) exp(-coef) lower .95 upper .95
## start_year     1.001     0.9989    0.9994     1.003
##
## Concordance= 0.5  (se = 0.012 )
## Likelihood ratio test= 1.73  on 1 df,   p=0.2
## Wald test            = 1.73  on 1 df,   p=0.2
## Score (logrank) test = 1.73  on 1 df,   p=0.2
```

The p-value for start__year is 0.188. Under a 0.05 and 0.10 significance level, the covariate start __year is not statistically significant on survival times.

III. Testing Proportional Hazard Assumptions

Log-log Graphs of Covariates

We first used the log-log graph technique to check whether the proportional hazards assumptions are met for our important covariates. A log-log curve is a transformation of an estimated survival curve. A log-log plot is a way to visually see if the assumptions are violated. If the lines are parallel, the assumptions are met. If the proportional hazards assumptions are met, this means that the relative hazard remains constant over time with different predictor or covariate levels.



Based on the log-log plot, the proportional hazards assumptions are violated since the lines are not parallel. However, the log-log plot does not work for continuous variables, and is clearly hard to read. As such, we continue looking at other tests.

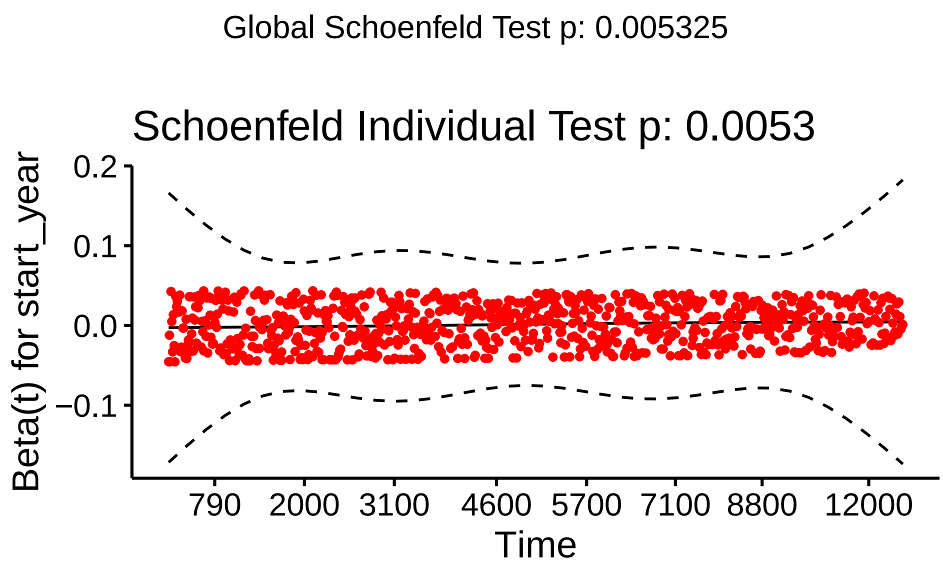
Testing Proportional Hazards with Coxzph

Conducting a statistical test using the fitted model above. The null hypothesis for this test is that the proportional hazards assumption is met. Thus with a .0053 p-value falling under our significance threshold of .05, we reject the null hypothesis and conclude that the proportional hazards assumption is not met.

```
test.ph <- cox.zph(res.cox2)
test.ph
```

```
##             chisq df      p
## start_year  7.77  1 0.0053
## GLOBAL      7.77  1 0.0053
```

```
## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values
```



The density of the residual plots seem to mostly center around a horizontal line at the Y=0 and the black line of best fit looks to reflect this as well. As such, the residuals seem to be random and constant with time. Because of this, we assume that the proportional hazards assumption is met for start_year.

IV. Further Analysis: Competing Risks Model

For a competing risks model, we assume independence of the specific causes for leaving. We also assume no recurrence of leaving the office (i.e. the senator in office leaving the office twice in one term). First, we code new status columns to censor the other causes of leaving office for each cause.

```
senators = senators %>%
  mutate(status_death = ifelse(reason=="Death", 1, 0),
         status_retire = ifelse(reason=="Retirement", 1, 0),
         status_resign = ifelse(reason=="Resignation", 1, 0))
```

Using these new status columns, we build separate models for each reason for leaving.

```
res.cox.death <- coxph(Surv(diff_days, status_death) ~ start_year,
                       data = senators)
summary(res.cox.death)
```

```
## Call:
## coxph(formula = Surv(diff_days, status_death) ~ start_year, data = senators)
##
##   n= 930, number of events= 494
##
##                  coef exp(coef)  se(coef)      z Pr(>|z|)
## start_year -0.008337  0.991697  0.001144 -7.285 3.21e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##            exp(coef) exp(-coef) lower .95 upper .95
## start_year    0.9917      1.008    0.9895    0.9939
```

```
##
## Concordance= 0.608  (se = 0.015 )
## Likelihood ratio test= 54.53  on 1 df,   p=2e-13
## Wald test            = 53.07  on 1 df,   p=3e-13
## Score (logrank) test = 54.33  on 1 df,   p=2e-13
```

Here, we see that our likelihood ratio's p-value of 1*10^(-12) is below 0.05, which means that for cause being death and controlling for start year, the model is statistically significant under a .05 significance level.

```
res.cox.retire <- coxph(Surv(diff_days, status_retire) ~ start_year,
                       data = senators)
summary(res.cox.retire)
```

```
## Call:
## coxph(formula = Surv(diff_days, status_retire) ~ start_year,
##     data = senators)
##
##   n= 930, number of events= 136
##
##                 coef exp(coef) se(coef)     z Pr(>|z|)
## start_year 0.067498  1.069828 0.006004 11.24   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##            exp(coef) exp(-coef) lower .95 upper .95
## start_year      1.07     0.9347     1.057     1.082
##
## Concordance= 0.878  (se = 0.01 )
## Likelihood ratio test= 320  on 1 df,   p=<2e-16
## Wald test            = 126.4  on 1 df,   p=<2e-16
## Score (logrank) test = 230.6  on 1 df,   p=<2e-16
```

Again, we see that our likelihood ratio's p-value of <2e-16 is below 0.05, which means that for cause being retirement and controlling for start year, the model is statistically significant under a .05 significance level.

```
res.cox.resign <- coxph(Surv(diff_days, status_resign) ~ start_year,
                       data = senators)
summary(res.cox.resign)
```

```
## Call:
## coxph(formula = Surv(diff_days, status_resign) ~ start_year,
##     data = senators)
##
##   n= 930, number of events= 201
##
##                 coef exp(coef) se(coef)     z Pr(>|z|)
## start_year 0.002721  1.002724 0.001758 1.548    0.122
##
##            exp(coef) exp(-coef) lower .95 upper .95
## start_year     1.003     0.9973    0.9993     1.006
##
## Concordance= 0.489  (se = 0.028 )
```

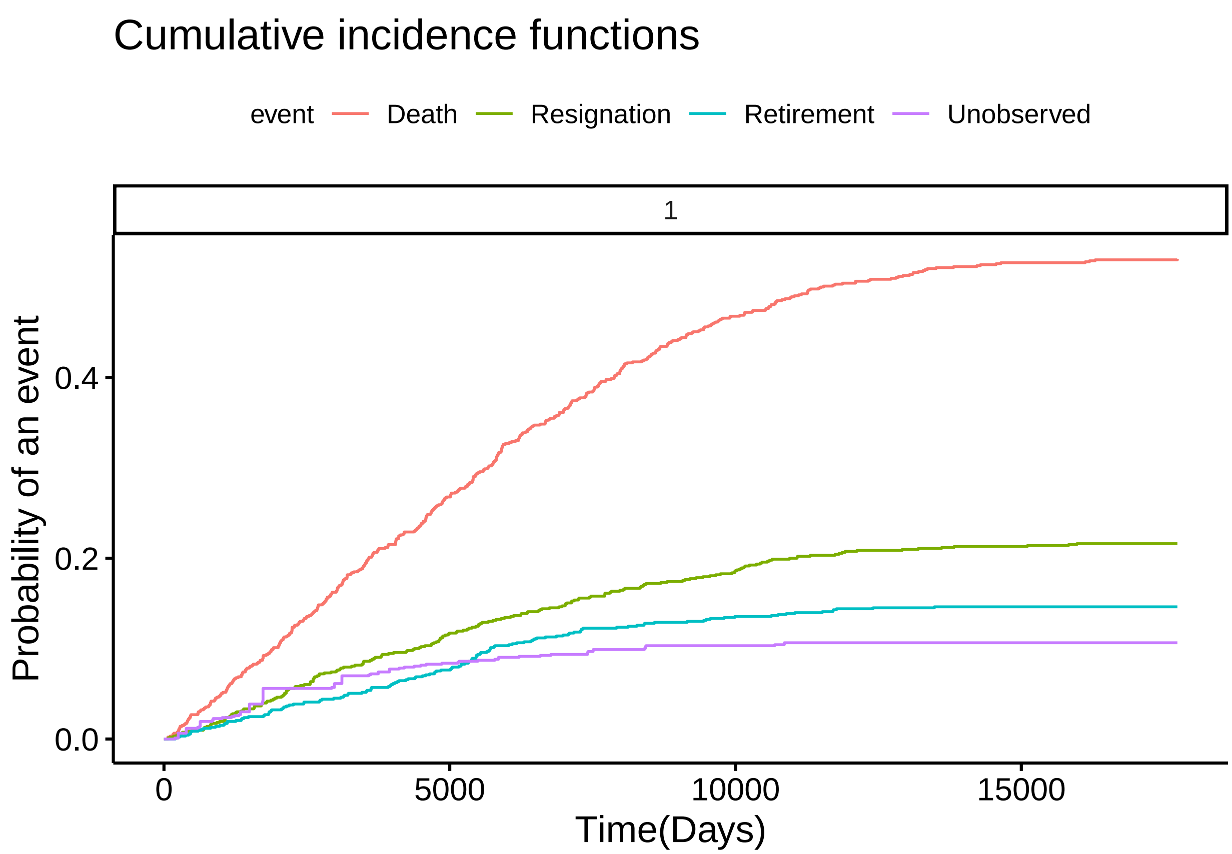
```
## Likelihood ratio test= 2.41  on 1 df,   p=0.1
## Wald test            = 2.4  on 1 df,   p=0.1
## Score (logrank) test = 2.4  on 1 df,   p=0.1
```

Here, we see that the likelihood's p-value is 0.08 which is greater than 0.05, so it is not statistically significant under a .05 significance level. The start year covariate's coefficient p-value also drops below 0.05 and is not statistically significant.

Cumulative Incidence Curve

Here we get the cumulative incidence for our data and plot the competing risks. This shows the probability of our competing events by time. Shown below, death leads as the most probable reason for Canadian senators to leave office followed by resignation and retirement.

```
cif <- cuminc(ftime=senators$diff_days, fstatus=senators$reason)
ggcompetingrisks(cif, xlab="Time(Days)")
```



Conclusion

We summarize the cause-specific models below with their hazard ratios and 95% confidence intervals.

Death as Cause

Characteristic	HR	95% CI	p-value
start_year	0.99	0.99, 0.99	<0.001

Retirement as Cause

Characteristic	HR	95% CI	p-value
start_year	1.07	1.06, 1.08	<0.001

Resignation as Cause

Characteristic	HR	95% CI	p-value
start_year	1.00	1.00, 1.01	0.12

The start year does not show a statistically significant effect on the resignation-specific model. However, the start year shows a statistically significant decrease for the death-specific model and a statistically significant increase for the retirement-specific model. This may be due to a constitutional amendment in 1965 that required Senators appointed after June 1, 1965 to retire at age 75. Though Senators appointed before that may still hold office for life, we may see a resulting time effect for senators pre-1965 and post-1965. This can be observed in conjunction with the comparison between the time effect on the death model and the retirement model. For death, time has a positive effect on survival and a negative effect on retirement. This is likely due to the fact that the older that the senator gets, the closer they are to retirement due to the mandatory retirement policy.

We can further consider stratifying on pre-1965 and post-1965 start year to test this.

V. Further Analysis: Stratifying on Pre-1965 and Post-1965 Start Year

Here, we are trying to observe the effect of the 1965 policy on the survival rates and how it influenced our model.

```
senators.new = senators %>%
  mutate(start1965 = case_when(start_year<=1965 ~ "Pre",
                               start_year>1965 ~ "Post"))

res.cox.death1965pre <- coxph(
  Surv(diff_days[start1965=="Pre"], status_death[start1965=="Pre"]) ~
    start_year[start1965=="Pre"], data = senators.new)
summary(res.cox.death1965pre)
```

```
## Call:
## coxph(formula = Surv(diff_days[start1965 == "Pre"], status_death[start1965 ==
##     "Pre"]) ~ start_year[start1965 == "Pre"], data = senators.new)
##
```

```
##   n= 599, number of events= 454
##
##                                       coef exp(coef)  se(coef)      z Pr(>|z|)
## start_year[start1965 == "Pre"] -0.000655  0.999345  0.001540 -0.425    0.671
##
##                                  exp(coef) exp(-coef) lower .95 upper .95
## start_year[start1965 == "Pre"]    0.9993      1.001    0.9963     1.002
##
## Concordance= 0.505  (se = 0.016 )
## Likelihood ratio test= 0.18  on 1 df,   p=0.7
## Wald test            = 0.18  on 1 df,   p=0.7
## Score (logrank) test = 0.18  on 1 df,   p=0.7
```

```
res.cox.death1965post <- coxph(
  Surv(diff_days[start1965=="Post"], status_death[start1965=="Post"]) ~
    start_year[start1965=="Post"], data = senators.new)
summary(res.cox.death1965post)
```

```
## Call:
## coxph(formula = Surv(diff_days[start1965 == "Post"], status_death[start1965 ==
##     "Post"]) ~ start_year[start1965 == "Post"], data = senators.new)
##
##   n= 331, number of events= 40
##
##                                       coef exp(coef) se(coef)      z Pr(>|z|)
## start_year[start1965 == "Post"] -0.03610   0.96454  0.01485 -2.432    0.015 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                   exp(coef) exp(-coef) lower .95 upper .95
## start_year[start1965 == "Post"]    0.9645      1.037    0.9369     0.993
##
## Concordance= 0.607  (se = 0.052 )
## Likelihood ratio test= 6.23  on 1 df,   p=0.01
## Wald test            = 5.91  on 1 df,   p=0.02
## Score (logrank) test = 6.15  on 1 df,   p=0.01
```

```
res.cox.retire1965pre <- coxph(
  Surv(diff_days[start1965=="Pre"], status_retire[start1965=="Pre"]) ~
    start_year[start1965=="Pre"], data = senators.new)
summary(res.cox.retire1965pre)
```

```
## Call:
## coxph(formula = Surv(diff_days[start1965 == "Pre"], status_retire[start1965 ==
##     "Pre"]) ~ start_year[start1965 == "Pre"], data = senators.new)
##
##   n= 599, number of events= 0
##
##                                  coef exp(coef) se(coef)  z Pr(>|z|)
## start_year[start1965 == "Pre"]   NA        NA        0 NA       NA
##
##                                  exp(coef) exp(-coef) lower .95 upper .95
## start_year[start1965 == "Pre"]        NA         NA        NA        NA
```

```
##
## Concordance= NA  (se = NA )
## Likelihood ratio test= 0  on 0 df,   p=1
## Wald test            = 0  on 0 df,   p=1
## Score (logrank) test = 0  on 0 df,   p=1
```

As seen by our results, none of the senators retired before the law came into effect.

```
res.cox.retire1965post <- coxph(
  Surv(diff_days[start1965=="Post"], status_retire[start1965=="Post"]) ~
    start_year[start1965=="Post"], data = senators.new)
summary(res.cox.retire1965post)
```

```
## Call:
## coxph(formula = Surv(diff_days[start1965 == "Post"], status_retire[start1965 ==
##     "Post"]) ~ start_year[start1965 == "Post"], data = senators.new)
##
##   n= 331, number of events= 136
##
##                                    coef exp(coef) se(coef)     z Pr(>|z|)
## start_year[start1965 == "Post"] 0.02915   1.02958  0.00841 3.467 0.000527 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                 exp(coef) exp(-coef) lower .95 upper .95
## start_year[start1965 == "Post"]      1.03     0.9713     1.013     1.047
##
## Concordance= 0.595  (se = 0.027 )
## Likelihood ratio test= 12.48  on 1 df,   p=4e-04
## Wald test            = 12.02  on 1 df,   p=5e-04
## Score (logrank) test = 12.29  on 1 df,   p=5e-04
```

Conclusions

As seen by our result summaries below, the pre-1965 model on death did not have a significant time effect, as the p-value is 0.7. However, the post-1965 death model showed a significant time effect with a p-value of 0.015, where the start year decreases the days in office for senators. The pre-1965 retirement model was omitted as there were no events, meaning there were no ended terms due to retirement before 1965. This is consistent with the introduction of the 1965 mandatory retirement policy. Finally, the post-1965 retirement model showed that the start year caused a statistically significant increase in days in office for senators under a 0.05 significance level, with a p-value less than <0.001.

Compared to before we split the data into pre-1965 and post-1965, the hazard rate for post-1965 death decreased from 0.99 to 0.96. This means the recalculated increase in survival rate for senators who left due to death is lower for post-1965. This may also be influenced by the improvements of technology resulting in increased life expectancy, as the drastic improvements to life expectancy and health has slowed. Meanwhile the post-1965 retirement hazard rate, compared to before the data split, changed from 1.08 to 1.03. This means the recalculated decrease in survival rate for senators who left due to retirement is lower for post-1965. We can conclude that there is a significant difference when considering the 1965 policy's effect on our model, especially on retirement. The lower decrease in survival rate may be due to the previous data including a number of senators who was not influenced by the 1965 policy.

Death as Cause

Characteristic	HR	95% CI	p-value
start_year[start1965 == “Pre”]	1.00	1.00, 1.00	0.7

Characteristic	HR	95% CI	p-value
start_year[start1965 == “Post”]	0.96	0.94, 0.99	0.015

Retirement as Cause

Characteristic	HR	95% CI	p-value
start_year[start1965 == “Post”]	1.03	1.01, 1.05	<0.001

VI. References

Cunningham, R., & Wehrle, D. (1994). A Note on the Average Age of Senators Since Confederation. Canadian Parliamentary Review. Retrieved December 11, 2022, from http://www.revparl.ca/english/issue.asp?param=151&art=1024.

Davidson-Pilon, (2019). lifelines: survival analysis in Python. Journal of Open Source Software, 4(40), 1317, https://doi.org/10.21105/joss.01317.