

STAA57 Project

2025-01-04

1. Description of dataset

The dataset present in this report contains 9,856 rows and 8 records. Each record holds information corresponding to a subgroup of part time employees in Canada.

- Year: Year of observation (2006)
- Geography: The region specific to the observations
- Immig.: Categorical variable that holds the immigration status of the observed group including:
 - “Born in Canada” (this group is considered Non-Immigrants)
 - “Total Landed immigrants”(this is the sum of all immigrants in a subgroup)
 - “Very recent immigrants, 5 years or less”
 - “Recent immigrants 5+ years”
 - “Established immigrants, 10+ years”
 - “Recent immigrants, 5+ to 10 years”
 - “Non-landed immigrants”
- Reason : Categorical variable that holds the reasons that part time workers are working part-time:
 - “Part-time employment, all reasons”
 - “Own illness”
 - “Caring for children”
 - “Other personal or family responsibilities”
 - “Going to school”
 - “Personal preference”
 - “Other voluntary”
 - “Business conditions, did not look for full-time work in last month”
 - “Could not find full-time work, did not look for full-time work in last month”
 - “Business conditions, looked for full-time work in last month”
- Age group: The age group of a subgroup ranging from young teen to elderly.
- Men: The amount of men in a part-time employee subgroup
- Women: The total number of women in a part-time employee subgroup
- Both sexes: The sum of the previous two mentioned records.

1.2 Summary

Year	Geography	Immig.	Reason
Min. :2006	Length:9856	Length:9856	Length:9856
1st Qu.:2009	Class :character	Class :character	Class :character
Median :2012	Mode :character	Mode :character	Mode :character
Mean :2012			
3rd Qu.:2016			
Max. :2019			
Age.group	Both.sexes	Men	Women
Length:9856	Min. : 0.00	Min. : 0.00	Min. : 0.00
Class :character	1st Qu.: 0.00	1st Qu.: 0.00	1st Qu.: 0.00
Mode :character	Median : 7.20	Median : 2.30	Median : 4.80
	Mean : 62.63	Mean : 20.91	Mean : 41.51
	3rd Qu.: 35.00	3rd Qu.: 11.10	3rd Qu.: 22.80
	Max. :3610.60	Max. :1288.90	Max. :2321.70

2. Background of dataset

This dataset is sourced by Statistics Canada. Since January 1997 they have surveyed workers who work less than 30 hours a week and hence are classified as part-time workers, asking if working 30 hours a week is sufficient for them or not. Depending on their response the main reason for working part-time is derived. For those who claim the work they have is adequate, the reason for not wanting to work full-time is collected. And for those who claim that they wish to work full time, the reason for working part-time is collected. Along with the reason for working part-time, sex, age, location, year, and immigration status is also collected. Involuntary part-time rate can also be calculated by dividing the number of persons whose response was business conditions or could not find work with 30 or more hours by the total number of persons working part-time at their main or only job. **Analyzing this dataset can help us discern broader economic trends, job availability and workforce participation patterns.**

3. Overall Research question

I aim to observe and analyze some of the factors that affect part-time employment trends in Canada, particularly peoples satisfaction with working part-time and their reasons for working part-time.

The factors I intend to investigate include:

- Age
- Sex
- Immigration status

4. Tables

4.1 Immigrant vs Non-Immigrant proportion of involuntary part-time work (per year)

Immigrant vs Non-Immigrant
proportion of involuntary part-time
workers

Year	Immigrant	Non-Immigrant
2006	23.4	20.4
2007	22.9	18.9
2008	23.4	19.3
2009	29.9	22.5
2010	29.7	22.4
2011	27.8	22.1
2012	28.7	21.9
2013	28.8	21.9
2014	29.3	22.3
2015	27.5	21.5
2016	24.2	19.9
2017	23.0	19.2
2018	21.4	16.9
2019	19.3	15.3

The table displays the involuntary part-time rate of Immigrants and Non-Immigrants (eg: total involuntary immigrants/total part-time working immigrants) The disparity between Immigrants and Non-Immigrant part-time workers is clear, every year a higher percentage of Immigrants work part-time involuntarily, compared to the Non-Immigrant. The range of the this disparity is no more than 10% for each year.

4.2 Most common reason, Men vs Women (per year)

Most common reason for working part-time, Men vs Women

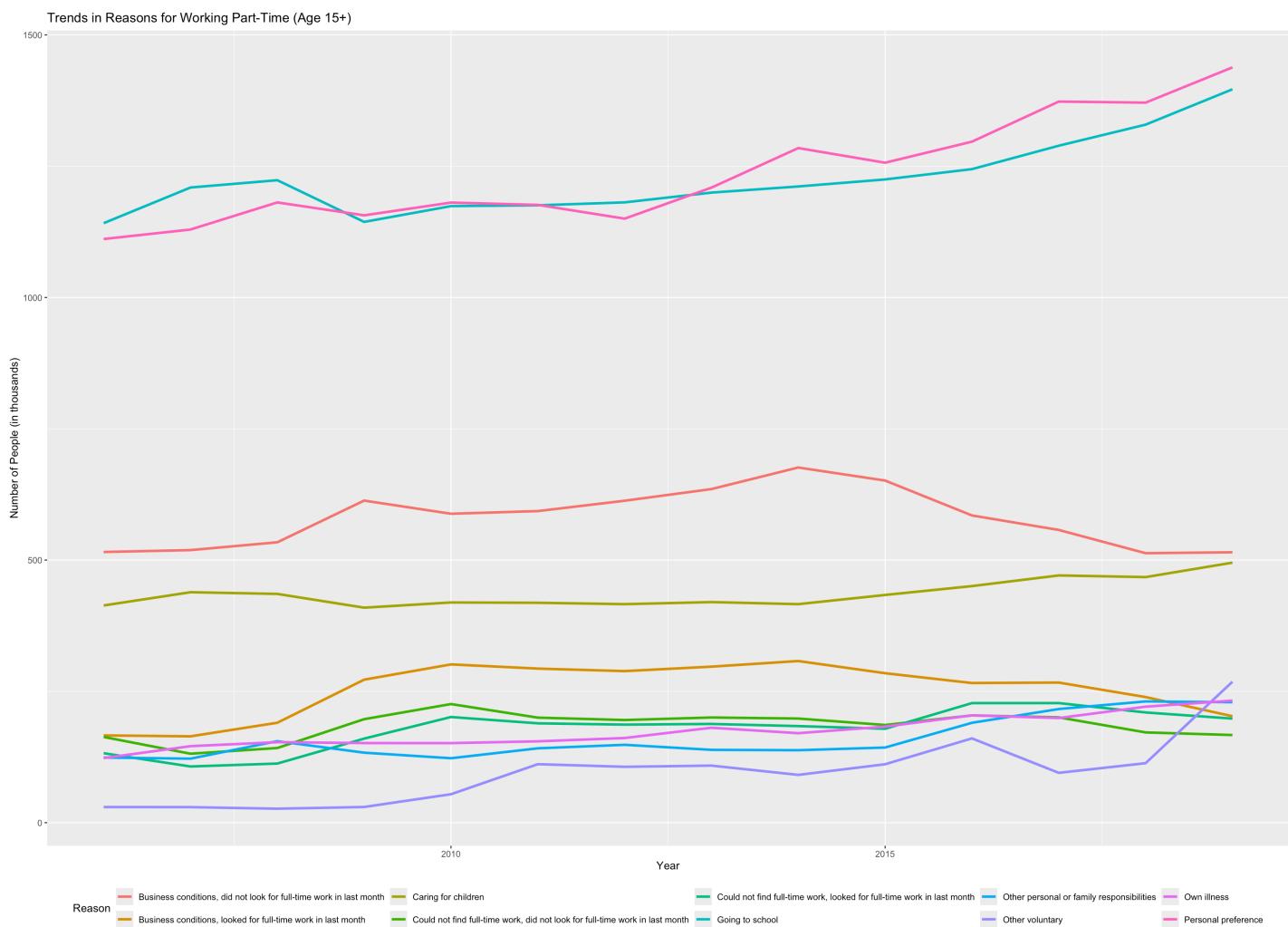
Year	Most Common Reason (Men)	Most Common Reason (Women)
2006	Going to school	Personal preference
2007	Going to school	Personal preference
2008	Going to school	Personal preference
2009	Going to school	Personal preference
2010	Going to school	Personal preference
2011	Going to school	Personal preference
2012	Going to school	Personal preference
2013	Going to school	Personal preference
2014	Going to school	Personal preference
2015	Going to school	Personal preference
2016	Going to school	Personal preference
2017	Going to school	Personal preference

Year	Most Common Reason (Men)	Most Common Reason (Women)
2018	Going to school	Personal preference
2019	Going to school	Personal preference

The table records the most common reason for working part-time for each sex per year, the trend of men working part-time due to school and women working part-time as per their own personal preference, may be a result of underlying traditional gender roles or differing opportunities for each sex, but more information that goes beyond the scope of this research would be required to make any definite conclusions

5. Graphs

5.1 Trends in reasons for working part-time per year, all collected ages.

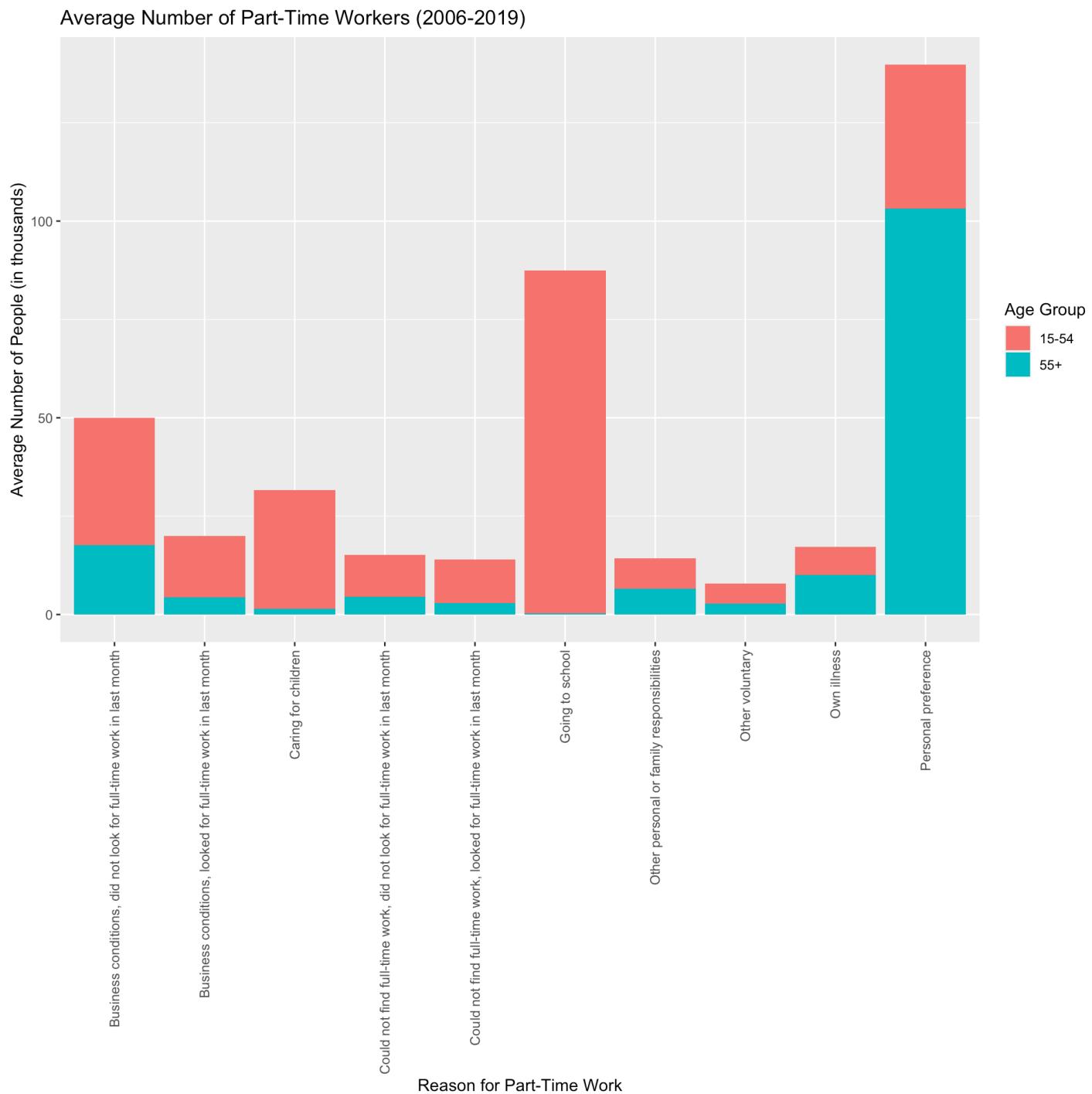


Across the years, economic and societal trends may vary, as a result, this graph displays that the reasons for working part-time as both an Immigrant and Non-Immigrant is ever-changing.

The graph dictates that overall the two most common motives for part-time work is preference and school.

Other trends can be seen as well such as the “caring for children” reason increasing which may relate to an increase in birth-rate

5.2 Stacked bar plot of Average number of part-time workers by Age-group and Reason (Average of 2006-2019)



The graph shows the trend of reason between age groups. It appears as though the higher age group has less people working part-time on average. Unsurprisingly something like “going to school” is not a common reason for someone older. It also appears that a higher proportion of elderly compared to young in the illness reason than elderly compared to young in any other reason, you can notice this as the illness bar is more covered by blue more so than the other reasons’ bars

6. Hypothesis Testing

I would like to investigate whether there is significant difference in the average proportion of elderly who claim they work part time due to illness and the average proportion of young people who claim the same.

H_0 : Older workers are less or equally likely to report illness than younger workers ($\mu_o \leq \mu_y$)

H_a : Older workers are more likely to report illnesses than younger workers ($\mu_o > \mu_y$)

```
##  
## Welch Two Sample t-test  
##  
## data: prop_old and prop_young  
## t = 22.686, df = 22.787, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 0.03176543      Inf  
## sample estimates:  
## mean of x  mean of y  
## 0.06395144 0.02958903
```

A Welch two-sample t-test was conducted to compare the proportion of part-time workers citing “own illness” as a reason for working part-time between young (ages 15–54) and elderly (ages 55+). Samples were obtained by creating a vector of the proportion of “Own illness” part-time workers for each year, and doing this for both age groups. The results showed a statistically significant difference ($t = 22.686$, $df = 22.787$, $p < 2.2e-16$), with older workers having a significantly higher average proportion ($M = 6.40\%$) compared to younger workers ($M = 2.96\%$). The 95% confidence interval for the difference in means was higher than 3.18% to. These findings support the hypothesis that older part-time workers are more likely to cite illness as their primary reason for part-time employment, and thus we reject the null hypothesis.

7. Summary

Some of the findings I have gathered are:

- For men the most common reason for part-time work is school related, while for women it is of their own preference
- In recent years (till 2019) preference is the most common reason to work part-time
- Before that it was school
- Immigrants working part-time are less likely to be working part-time of their own volition than Non-Immigrants working part-time
- The disparity in the proportion working part-time involuntarily between Immigrants and Non-Immigrants is no higher than 10%
- T
- There is a lesser amount of elderly(55+) working part-time compared to non elderly
- There is higher average proportion of elderly working part-time due to illness compared to non-elderly working part time due to illness.

Overall I think this data-set and report highlights the patterns and trends that occur in the reasons for working part-time as a result of differing factors such as the year, sex, age, immigration status. I think some apparent problems can be observed such as one the more common reasons being “Going to school”, perhaps people find it difficult to maintain payment of their school tuition without working part-time. It becomes more worrying when

considering that some of these people aren't adults. The difference in the most common reason between men and women may be hinting at an underlying issue. The amount of people working part-time involuntarily as a result of illness shows that Canada is lacking in accommodation for those with health issues.

8. Appendix

```
library(tidyverse)
library(ggplot2)
library(grid)
library(knitr)
library(kableExtra)
library(stringr)

data = read.csv("~/Downloads/g0811_05.csv")
data <- data %>% mutate(across(where(is.character), ~ str_squish(.)))
# 1.2 Summary
summary(data)
```

```

#4.1
involuntary_reasons <- c(
  "Business conditions, did not look for full-time work in last month",
  "Could not find full-time work, did not look for full-time work in last month",
  "Business conditions, looked for full-time work in last month"
)

d_filtered1 <- data %>%
  filter(
    Age.group == "15+",
    Immig. != "Total",
    Reason != "Part-time employment, all reasons"
  ) %>%
  mutate(
    Reason = str_trim(Reason),
    Immig. = str_trim(Immig.),
    Immigrant = if_else(Immig. == "Born in Canada", "Non-Immigrant", "Immigrant"),
    Involuntary = if_else(Reason %in% involuntary_reasons, Both.sexes, 0)
  )
summary_table <- d_filtered1 %>%
  group_by(Year, Immigrant) %>%
  summarise(
    Total_PT = sum(Both.sexes, na.rm = TRUE),
    Involuntary = sum(Involuntary, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  mutate(Percent_Involuntary = round(100 * Involuntary / Total_PT, 1)) %>%
  select(Year, Immigrant, Percent_Involuntary)
summary_table <- summary_table %>%
  pivot_wider(names_from = Immigrant, values_from = Percent_Involuntary)

summary_table %>% kbl(caption = "Immigrant vs Non-Immigrant proportion of involuntary pa
rt-time workers") %>%
  kable_classic(full_width = F, html_font = "Cambria")

```

```

#4.2
d_filtered1 <- data %>%
  filter(
    Age.group == "15+",
    Immig. != "Total",
    Reason != "Part-time employment, all reasons"
  ) %>%
  mutate(
    Reason = str_trim(Reason),
    Immig. = str_trim(Immig.),
    Geography = str_trim(Geography)
  )
gender_reasons <- d_filtered1 %>% select(Year,Reason,Men,Women) %>% pivot_longer( cols =
c(Men,Women), names_to = "Gender",
values_to ="C
ount"
)
gender_reasons <- gender_reasons %>%
  group_by(Year, Gender, Reason) %>%
  summarise(Total = sum(Count, na.rm = TRUE), .groups = "drop")

gender_reasons <- gender_reasons %>%
  group_by(Year, Gender) %>%
  slice_max(order_by = Total, n = 1, with_ties = FALSE) %>%
  ungroup()
gender_reasons <- gender_reasons %>%
  select(Year, Gender, Reason) %>%
  pivot_wider(names_from = Gender, values_from = Reason) %>%
  rename(
    `Most Common Reason (Men)` = Men,
    `Most Common Reason (Women)` = Women
  )

gender_reasons %>% kbl(caption = "Most common reason for working part-time, Men vs Wome
n") %>%
  kable_classic(full_width = F, html_font = "Cambria")

```

```

#5.1
d_filtered <- data %>%
  filter(
    Age.group == "15+",
    Immig. != "Total",
    Reason != "Part-time employment, all reasons"
  )

d_summary <- d_filtered %>%
  group_by(Year, Reason) %>%
  summarise(Total = sum(Both.sexes, na.rm = TRUE), .groups = 'drop')

ggplot(d_summary, aes(x = Year, y = Total, color = Reason)) +
  geom_line(size = 1.2) +
  labs(
    title = "Trends in Reasons for Working Part-Time (Age 15+)",
    x = "Year",
    y = "Number of People (in thousands)",
    color = "Reason") + theme(legend.position = "bottom")

```

```

#5.2
d_filtered3 <- data %>%
  filter(Reason != "Part-time employment, all reasons") %>%
  filter(Age.group != "15+") %>%
  mutate(Reason = str_trim(Reason))

d_filtered3 <- d_filtered3 %>%
  mutate(AgeCategory = case_when(
    Age.group %in% c("15-24", "25-54") ~ "15-54",
    Age.group == "55+" ~ "55+"
  ))

avg_table <- d_filtered3 %>%
  group_by(Reason, AgeCategory) %>%
  summarise(AvgCount = mean(Both.sexes, na.rm = TRUE), .groups = "drop")

ggplot(avg_table, aes(x = Reason, y = AvgCount, fill = AgeCategory)) +
  geom_bar(stat = "identity") +
  labs(
    title = "Average Number of Part-Time Workers (2006-2019)",
    x = "Reason for Part-Time Work",
    y = "Average Number of People (in thousands)",
    fill = "Age Group"
  ) + theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))

```

```

# 6.
set.seed(47)
d_filtered4 <- data %>%
  filter(Age.group %in% c("15-24", "25-54", "55+")) %>%
  mutate(
    AgeGroup = if_else(Age.group %in% c("15-24", "25-54"), "15-54", "55+")
  )

illness_prop <- d_filtered4 %>%
  filter(
    Geography == "Canada",
    Reason %in% c("Own illness", "Part-time employment, all reasons")
  ) %>%
  group_by(Year, AgeGroup, Reason) %>%
  summarise(Total = sum(Both.sexes, na.rm = TRUE), .groups = "drop") %>%
  pivot_wider(names_from = Reason, values_from = Total) %>%
  rename(
    Total_PT = "Part-time employment, all reasons",
    Illness = "Own illness"
  ) %>%
  mutate(Proportion = Illness / Total_PT)

prop_young <- illness_prop %>% filter(AgeGroup == "15-54") %>% pull(Proportion)
prop_old   <- illness_prop %>% filter(AgeGroup == "55+")   %>% pull(Proportion)

t.test(prop_old, prop_young, alternative = "greater")

```