Trends in Work-related Injury Frequency Rate in Canada*

Maritime operations have experienced a notable decline, contributing to an overall downward trend

Missy Zhang

16 April 2023

This paper aims to examine whether and how work-related injury rate has declined in Canada from 2008 to 2020. Using Statistics Canada data, it was found that there was an overall downward trend in the frequency of injuries over time, with the decline most pronounced in maritime operations. Particularly, it is closely related to the proportion of outdoor workers in the industry. In addition, this paper shares insights into injury prevention strategies for Canadian workplaces, and highlights the industries and worker populations that may be most vulnerable to workplace injuries.

1 Introduction

In recent years, work-related injuries have emerged as a critical issue in Canada, affecting workers across various industries. These injuries can have devastating physical and psychological impacts on workers and can also result in a significant economic burden on society. Despite a decrease in the number of injuries in recent years, there were still 37,024 work-related injuries reported in 2020 alone (Canada Employment and Social Development 2022). To address this issue and improve workplace safety, it is crucial to understand injury trends over time and across different sectors. This report aims to analyze work-related injury data and explore factors that contribute to work-related injuries. By identifying areas that require improvement and highlighting successful initiatives, this report aims to contribute to the ongoing efforts to create safer workplaces and reduce the incidence of work-related injuries.

This report utilizes data of injury statistics in Canada by industry sector from 2008 to 2020 obtained from the Open Government Data of Statistics Canada (Statistics Canada 2023) and

---

*Code and data are available at: https://github.com/MissyZhang/occupational_injury.

used programming language R (R Core Team 2020) to build models and analyze trends. The estimand of this report is work-related injury frequency rate, which is measured by injuries (disabling, fatal and minor) per one million hours worked. This paper delves into the overall trend of injury frequency rate in Canada by industry using multiple linear regression model. It was found that there was an overall downward trend in the frequency of injuries over time, with the decline most pronounced in maritime operations. Particularly, it is closely related to the proportion of outdoor workers in the industry.

The remainder of the paper is split into four sections. Section 2 explores the cleaned dataset and identifies key variables. Visualizations are presented to show correlations. Section 3 observes injury frequency rate has decreased over time before constructing the regression model. Section 4 presents model findings on injury frequency rate evolution. Section 5 discusses implications, solutions, and study limitations.

2 Data

2.1 Data source and methodology

The dataset provides a summary of occupational injury statistics and rates in the Canadian Federal Jurisdiction between 2008 and 2020, and it was sourced from the Open Government Portal of Statistics Canada. The data was collected from the Employer's Annual Hazardous Occurrence Report (EAHOR), which is submitted by federally regulated employers to the Labour Program. Under the Canada Labour Code, federal jurisdiction encompasses work and/or undertakings that fall under the legislative authority of the Parliament of Canada, including work deemed to be for the common good of at least two provinces and outside the exclusive authority of provincial legislatures Employment and Canada (2020). Industries under Federal Jurisdiction include interprovincial and international transportation, telecommunications and broadcasting, banks, postal services, feed, flour, seed and grain, miscellaneous industries, and first nation band councils and indigenous self-governments Employment and Canada (2021). One limitation of the survey methodology is that the it only captures incidents in industries under federal jurisdiction, which excludes provincially regulated workplaces and industries not classified as falling under federal jurisdiction. As a result, the dataset may not be representative of the entire Canadian workforce and could lead to an underestimation of the true incidence and severity of occupational injuries in Canada.

The original dataset contains 317 observations and 16 variables with all kinds of information about injury statistics and rates. This report wants to focus on exploring the trend of injury rate and investigating the possible factors that may affect injury rate. Thus, I will be interested in the variables "Year", "Industry", and "Injury frequency rate". R (R Core Team 2020), and R packages "tidyverse" (Wickham et al. 2019), "dplyr" (Wickham et al. 2021), and "kableExtra" (Zhu 2021) are utilized to create an extract of the cleaned dataset (Table 1).

Table 1: Extracting the first ten rows from the Injury data

Industry	Year	Injury Frequency Rate
Air Transportation	2020	39.7
Air Transportation	2019	36.5
Air Transportation	2018	39.8
Air Transportation	2017	45.4
Air Transportation	2016	65.3
Air Transportation	2015	54.0
Air Transportation	2014	55.1
Air Transportation	2013	49.2
Air Transportation	2012	72.2
Air Transportation	2011	80.2

Table 1 shows the first ten rows of the cleaned dataset. It contains 231 observations and 3 variables in total. The injury frequency rate is measured by all occupational injuries (disabling, fatal and minor) per one million hours worked in a given industry and year.

2.2 Data Visualization

In order to get further familiarized with the dataset and estimate the possible associations between the nation's injury rate and other factors, exploratory analysis is carried out by conducting data visualizations to observe whether the patterns between certain factors matches the generally expectations of the trend of injury rate in Canada. First, we will have a look through the overall trend of injury rate in Canada.
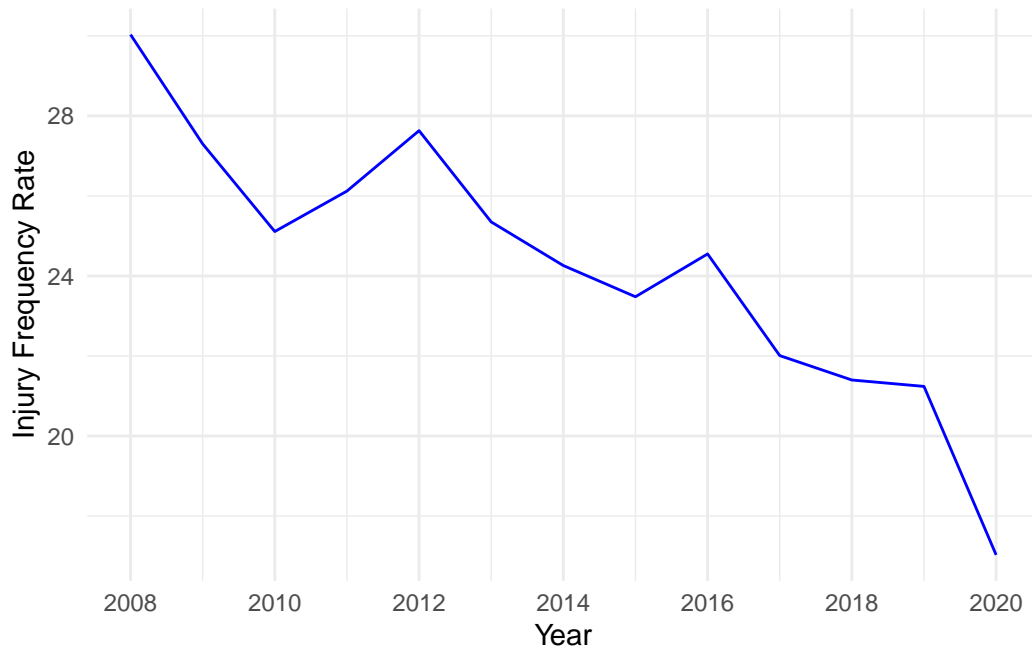
Figure 1: Work-related injury frequency rate by year in Canada from 2008 to 2020

As Figure 1 has demonstrated, Canada's overall injury frequency rate has fluctuated downward since 2008, indicating an improvement in the safety performance of federally regulated workplaces. However, there were two notable periods of increase in the injury frequency rate in 2010-2012 and 2015-2016. Despite these increases, the average injury frequency rate in Canada decreased from 30 in 2008 to 17 in 2020, indicating an overall positive trend in occupational safety.
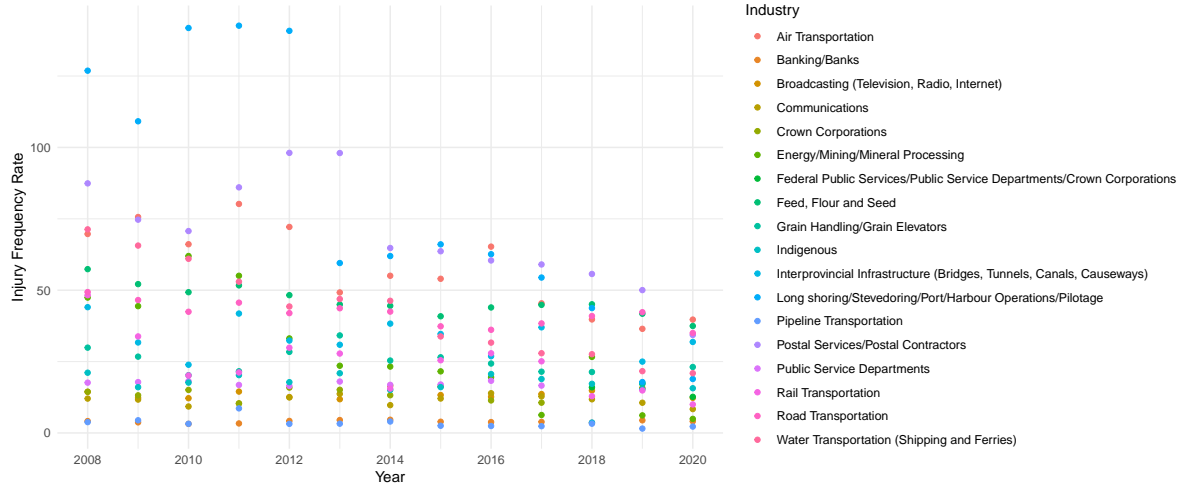
Figure 2: Work-related injury frequency rate by year and industry in Canada from 2008 to 2020

Figure 2 provides a breakdown of the data by different industries, revealing varying degrees of improvement in injury frequency rates over time. Notably, the Long shoring/Stevedoring/Port/Harbour Operations/Pilotage industry has shown the greatest decrease in injury frequency rate among all industries, as indicated by the blue dots on the figure. In contrast, Banking/Banks displays the smallest decrease in injury frequency rate, with a flatter slope in the figure. The differences in injury frequency rates among these industries can likely be attributed to variations in their work environments and tasks. For example, industries such as Long shoring/Stevedoring/Port/Harbour Operations/Pilotage may have implemented more rigorous safety protocols and procedures due to the inherently high-risk nature of their work, while Banking/Banks may not face the same physical hazards and thus may not have undergone the same level of safety improvements.

3 Model

By the exploratory analysis of the data, it was found that year and industry have some correlation with the injury frequency rate in Canada. The relationship seems to be linear as the figures generally demonstrate decreasing trends. Therefore, to further proceed the analysis and predict the future situation regarding injury rate, a multiple linear regression model will be constructed.

Prior to constructing the model, the dataset is split into the training set and the testing set with a proportion of 8:2. The training set is used to build the multiple linear regression models, and the testing set is used to test the accuracy and unbiasedness of the model. R package "tidymodels"(Kuhn and Wickham 2020) is used to split the dataset.

Two models are initially constructed based on the different possible factors that might have an effect on injury frequency rate in Canada. The first model has continuous variable year, categorical variable industry as its predictor variables, and the injury frequency rate as its response variable. I would also like to check whether the variables year and industry are implicitly related. Therefore, the second model has continuous variable year, categorical variable industry, and the interaction between them as its predictor variables, and the injury frequency rate as its response variable. To compare these two models to find out which is more accurate and is better at prediction, AIC and BIC tests are carried out, and R^2 is examined. Furthermore, the testing dataset is used to test how the two models perform when extra data is involved. RMSE is measured by the testing data to compare the prediction power of the two models. In Appendix Section A, these tests are performed and the test statistics of the two models are compared to select the model with better performance.

As a result, model 2 has a better performance than model 1, and is selected as the final model to model the behavior of Income inequality in Canada as time passed. The assumption check for the model is done in Appendix Section B.

The final model is displayed below:

$$Y_{ij} = \beta_0 + \beta_1 Year_i + \beta_2 Industry_j + \beta_3 Year_i Industry_j \tag{1}$$

In Model Equation 1:

- Y_{ij} is the injury frequency rate in i^{th} year and industry j.
- β_0 is the coefficient for intercept.
- β_1 is the coefficient for the continuous year variable.
- β_2 is the coefficient corresponding to industry j.
- β_3 is the coefficient for the interaction term between i^{th} year and industry j.
- The baseline of this model is year 0 and Air Transportation industry.

4 Results

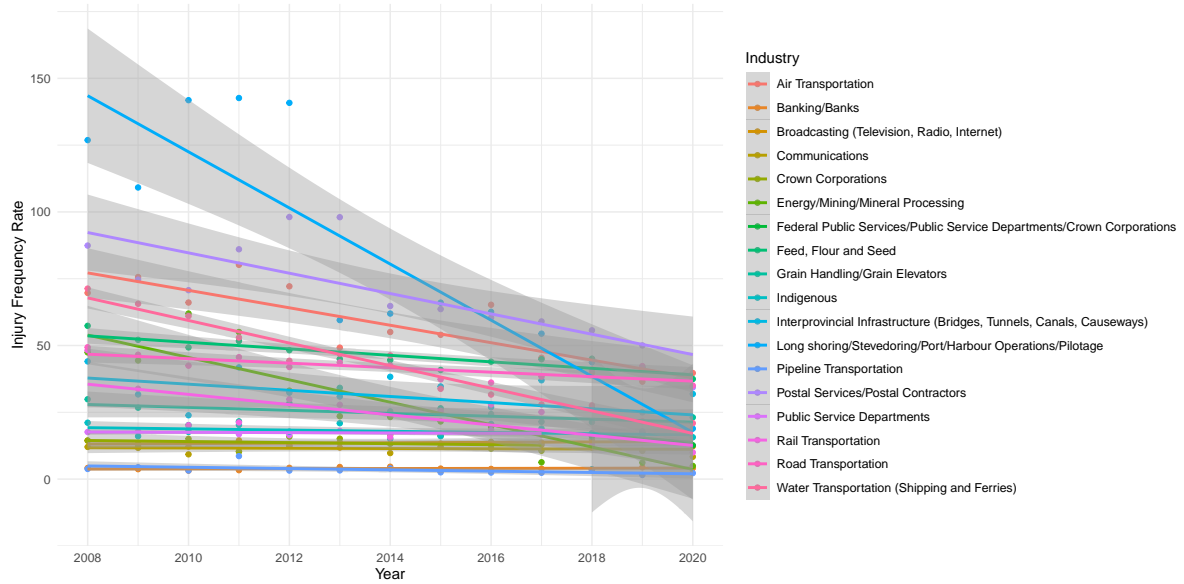Figure 3: Work-related injury frequency rate by year and industry in Canada from 2008 to 2020

Table 2: Model Coefficients and 95 percent Confidence Interval

	Coefficients	Confidence Interval Lower Bound	Confidence Interval Upper Bound
Intercept	6858.4	4365.0	9351.8
Banking/Banks	-6899.6	-10592.4	-3206.7
Broadcasting (Television, Radio, Internet)	-6868.3	-10464.3	-3272.2
Communications	-6545.3	-10288.0	-2802.6
Crown Corporations	-6444.2	-10560.6	-2327.7
Energy/Mining/Mineral Processing	1498.1	-1956.8	4953.0
Federal Public Services/Public Service Departments/Crown Corporations	-3481.4	-24732.5	17769.7
Feed, Flour and Seed	-4299.8	-7708.2	-891.3
Grain Handling/Grain Elevators	-5831.6	-9165.8	-2497.4
Indigenous	-6384.7	-9793.8	-2975.6
Interprovincial Infrastructure (Bridges, Tunnels, Canals, Causeways)	-6479.7	-10127.4	-2831.9
Long shoring/Stevedoring/Port/Harbour Operations/Pilotage	12705.9	9084.2	16327.5
Pipeline Transportation	-6303.4	-9723.9	-2882.9
Postal Services/Postal Contractors	-512.9	-4054.7	3028.9
Public Service Departments	-6583.9	-12730.4	-437.4
Rail Transportation	-3174.3	-6579.9	231.3
Road Transportation	-4923.1	-8358.9	-1487.3
Water Transportation (Shipping and Ferries)	1559.5	-2044.5	5163.6
Year	-3.4	-4.6	-2.1
Year: Banking/Banks	3.4	1.6	5.2
Year: Broadcasting (Television, Radio, Internet)	3.4	1.6	5.2
Year: Communications	3.2	1.4	5.1
Year: Crown Corporations	3.2	1.1	5.2
Year: Energy/Mining/Mineral Processing	-0.8	-2.5	1.0
Year: Federal Public Services/Public Service Departments/Crown Corporations	1.7	-8.8	12.2
Year: Feed, Flour and Seed	2.1	0.4	3.8
Year: Grain Handling/Grain Elevators	2.9	1.2	4.5
Year: Indigenous	3.1	1.5	4.8
Year: Interprovincial Infrastructure (Bridges, Tunnels, Canals, Causeways)	3.2	1.4	5.0
Year: Long shoring/Stevedoring/Port/Harbour Operations/Pilotage	-6.3	-8.1	-4.5
Year: Pipeline Transportation	3.1	1.4	4.8
Year: Postal Services/Postal Contractors	0.3	-1.5	2.0
Year: Public Service Departments	3.2	0.2	6.3
Year: Rail Transportation	1.6	-0.1	3.3
Year: Road Transportation	2.4	0.7	4.1
Year: Water Transportation (Shipping and Ferries)	-0.8	-2.6	1.0

5 Discussion

5.1 First discussion point

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Appendix

A Model Testing

R package "modelsummary"(Arel-Bundock 2022) helps to display the coefficients of the two models as well as the result of a series of tests to compare which model performs better.

R^2 measures how well the model explains its response variable's variation. If R^2 is low, it indicates that the model doesn't fit the data well. By Table 3, in Model 1, 84.5% of the variability is explained, and in Model 2, 87.4% of the variability is explained. Both models demonstrate a pretty high R^2 value.

AIC and BIC measure the prediction ability of the multiple linear regression model. AIC (Akaike's Information Criteria) focuses on how well the model fits unknown data, while BIC (Bayesian Information Criteria) focuses on the true model and favours simpler models (Kellen 2010). Lower AIC and BIC both indicate that the model has better prediction power. By Table 3, Model 2 has slightly lower AIC and BIC than Model 1, implying that Model 2 has more prediction power.

Table 3: Comparing Model 1 and Model 2's Statistics

	Model 1	Model 2
(Intercept)	3771.03	6858.39
	(487.75)	(1261.02)
IndustryBanking/Banks	-54.07	-6899.55
	(5.34)	(1867.60)
IndustryBroadcasting (Television, Radio, Internet)	-44.29	-6868.27
	(5.20)	(1818.67)
IndustryCommunications	-49.59	-6545.29
	(5.72)	(1892.81)
IndustryCrown Corporations	-48.75	-6444.17
	(5.19)	(2081.85)
IndustryEnergy/Mining/Mineral Processing	-29.67	1498.10
	(4.97)	(1747.28)
IndustryFederal Public Services/Public Service Departments/Crown Corporations	-34.97	-3481.40
	(7.75)	(10747.51)
IndustryFeed, Flour and Seed	-12.49	-4299.77
	(5.07)	(1723.77)
IndustryGrain Handling/Grain Elevators	-35.55	-5831.56
	(4.97)	(1686.24)
IndustryIndigenous	-41.71	-6384.71

	Model 1	Model 2
	(5.19)	(1724.11)
IndustryInterprovincial Infrastructure (Bridges, Tunnels, Canals, Causeways)	-26.92	-6479.69
	(5.19)	(1844.81)
IndustryLong shoring/Stevedoring/Port/Harbour Operations/Pilotage	23.57	12705.87
	(5.19)	(1831.60)
IndustryPipeline Transportation	-56.38	-6303.43
	(5.07)	(1729.89)
IndustryPostal Services/Postal Contractors	10.57	-512.93
	(5.33)	(1791.22)
IndustryPublic Service Departments	-47.68	-6583.88
	(5.75)	(3108.52)
IndustryRail Transportation	-35.16	-3174.31
	(4.97)	(1722.35)
IndustryRoad Transportation	-16.92	-4923.10
	(5.08)	(1737.60)
IndustryWater Transportation (Shipping and Ferries)	-17.59	1559.53
	(5.19)	(1822.71)
Year	-1.84	-3.38
	(0.24)	(0.63)
IndustryBanking/Banks $\times$ Year		3.40
		(0.93)
IndustryBroadcasting (Television, Radio, Internet) $\times$ Year		3.39
		(0.90)
IndustryCommunications $\times$ Year		3.23
		(0.94)
IndustryCrown Corporations $\times$ Year		3.18
		(1.03)
IndustryEnergy/Mining/Mineral Processing $\times$ Year		-0.76
		(0.87)
IndustryFederal Public Services/Public Service Departments/Crown Corporations $\times$ Year		1.71
		(5.32)
IndustryFeed, Flour and Seed $\times$ Year		2.13
		(0.86)
IndustryGrain Handling/Grain Elevators $\times$ Year		2.88
		(0.84)
IndustryIndigenous $\times$ Year		3.15
		(0.86)

	Model 1	Model 2
IndustryInterprovincial Infrastructure (Bridges, Tunnels, Canals, Causeways) × Year		3.21
		(0.92)
IndustryLong shoring/Stevedoring/Port/Harbour Operations/Pilotage × Year		-6.30
		(0.91)
IndustryPipeline Transportation × Year		3.10
		(0.86)
IndustryPostal Services/Postal Contractors × Year		0.26
		(0.89)
IndustryPublic Service Departments × Year		3.25
		(1.55)
IndustryRail Transportation × Year		1.56
		(0.86)
IndustryRoad Transportation × Year		2.44
		(0.86)
IndustryWater Transportation (Shipping and Ferries) × Year		-0.78
		(0.91)
Num.Obs.	174	174
R2	0.817	0.932
R2 Adj.	0.795	0.915
AIC	1366.6	1227.5
BIC	1429.8	1344.4
Log.Lik.	-	-576.767
	663.324	
RMSE	10.95	6.66

Table 4: Comparing RMSE between two models

Dataset	Model 1	Model 2
train	10.9	6.7
test	15.9	9.3

The RMSE measures how far the predicted values of the multiple linear regression model are from their actual values on average. The lower the RMSE, the better the model preforms regarding prediction. Based on Table 4, the RMSE for both two models are very similar for training and testing dataset, indicating that the dataset is unbiased and the two models are performing as expected. Moreover, Model 2 has lower RMSE than Model 1 for both training and testing dataset, indicating that Model 2 predicts data more accurately.

B Model Assumption Check
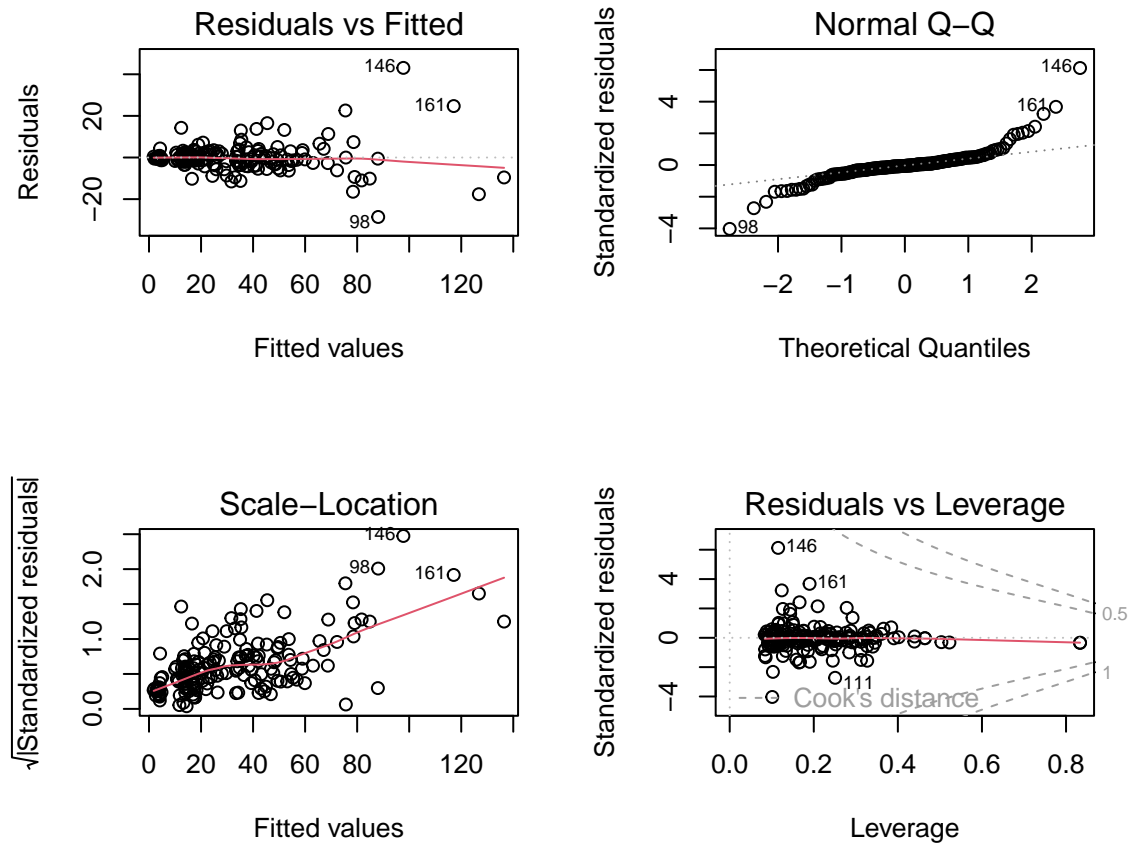
Figure 4: Checking the assumptions of linear model

Check for assumptions: Assumptions to the multiple linear regression model are checked to ensure that the model is valid for this dataset. Figure 4 shows the plots used to check for the assumption.

The Residuals vs Fitted plot checks for the linear relationship assumption. Since the red line is almost horizontal and there isn't any pattern, the model satisfies the linearity assumption. The Normal QQ plot checks for the residual normality assumption. Since almost all the dots are on the dashed line, the residuals follow a normal distribution. The Scale-Location plot checks for the homoscedasticity assumption. The red line is horizontal, and the dots are evenly scattered, indicating that the variance of the residuals is constant (Kassambara 2018).

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Canada Employment and Social Development. 2022. *2020 Annual Report - Occupational Injuries in the Canadian Federal Jurisdiction.* https://www.canada.ca/en/employment-social-development/services/health-safety/reports/2020-injuries.html.

Employment, and Social Development Canada. 2020. *2020 Annual Report - Occupational Injuries in the Canadian Federal Jurisdiction.* https://www.canada.ca/en/employment-social-development/services/health-safety/reports/2020-injuries.html.

———. 2021. *Industries Under Federal Jurisdiction.* https://www.canada.ca/en/employment-social-development/corporate/reports/labour-transition-binders/minister-labour-2021/industries-infographic.html.

Kassambara, Alboukadel. 2018. *Regression Model Diagnostics.* http://www.sthda.com/english/articles/39-regression-model-diagnostics/161-linear-regression-assumptions-and-diagnostics-in-r-essentials/.

Kellen, Dave. 2010. *Is There Any Reason to Prefer the AIC or BIC over the Other?* https://stats.stackexchange.com/questions/577/is-there-any-reason-to-prefer-the-aic-or-bic-over-the-other.

Kuhn, Max, and Hadley Wickham. 2020. *Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles.* https://www.tidymodels.org.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Statistics Canada. 2023. *Open Government Data of Statistics Canada.* https://www150.statcan.gc.ca/n1/en/type/data.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation.* https://dplyr.tidyverse.org, https://github.com/tidyverse/dplyr.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax.* http://haozhu233.github.io/kableExtra/, https://github.com/haozhu233/kableExtra.