Evolution of injury fruquency rate in Canada*

My subtitle if needed

First author Another author

4 April 2023

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

You can and should cross-reference sections and sub-sections. For instance, Section 2 and Section 5.1.

2 Data

Our data is of penguins (Figure 1).

*Code and data are available at: LINK.

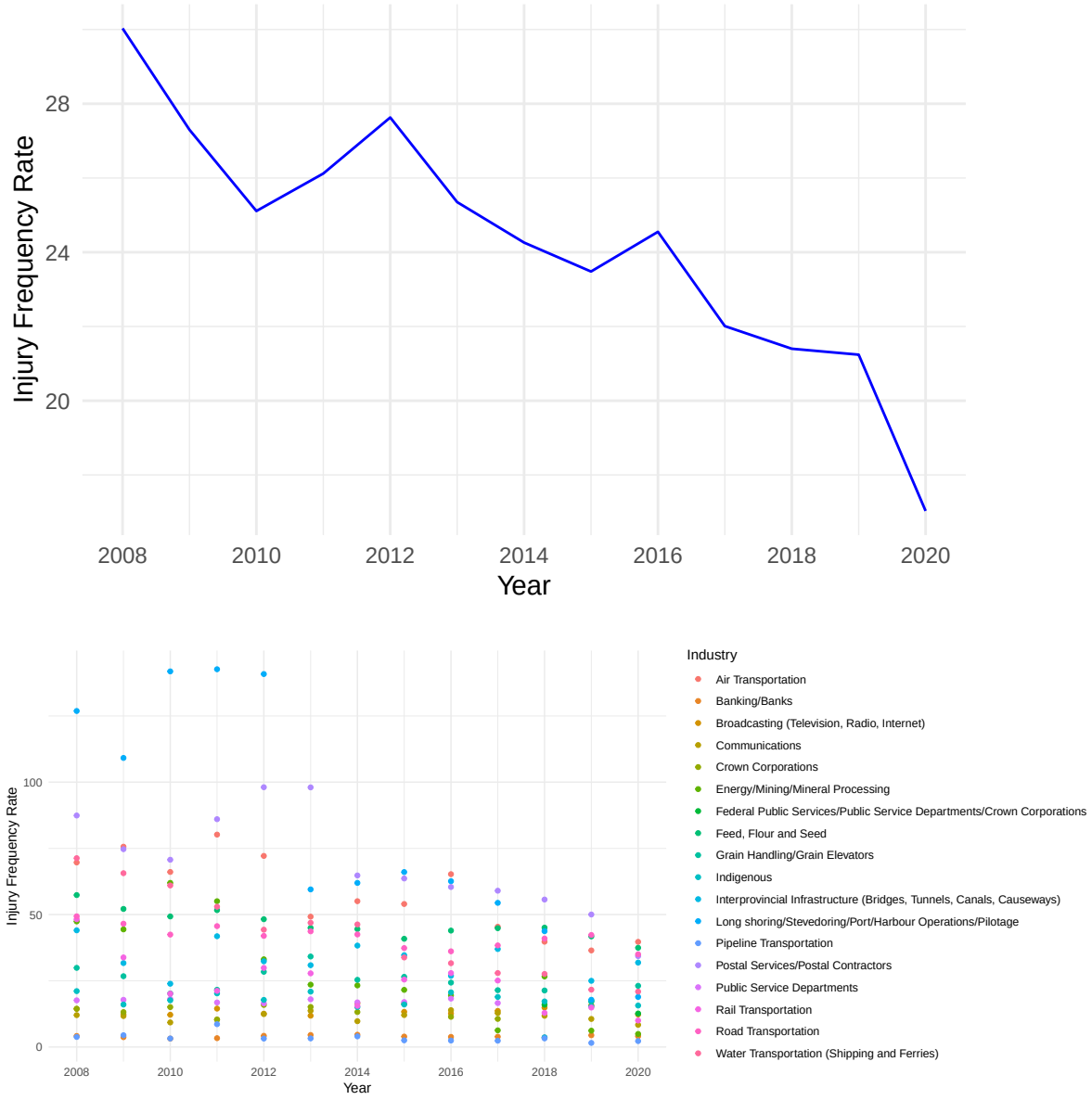


Figure 1: Bills of penguins

Talk more about it.

Also bills and their average (Figure 2). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work.)

Talk way more about it.

3 Model

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

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.

Here's a dumb example of how to use some references: In paper we run our analysis in `R` (R Core Team 2020). We also use the `tidyverse` which was written by If we were interested in baseball data then could be useful.

We can use maths by including latex between dollar signs, for instance θ.

4 Results

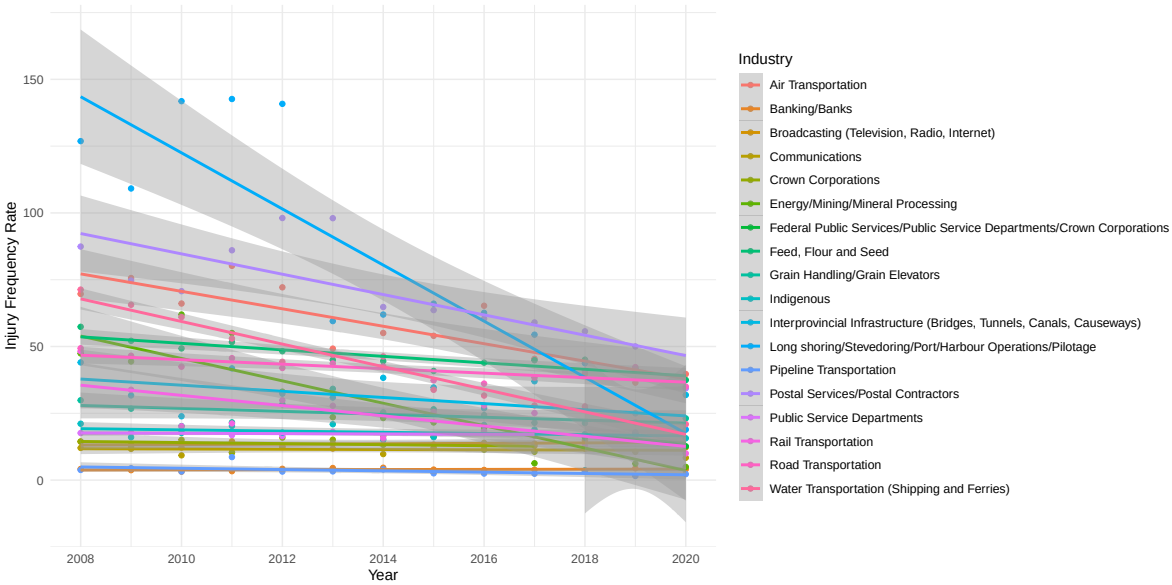


Figure 2: More bills of penguins

Table 1: Model Coefficients and 95 percent Confidence Interval

	Coefficients	Confidence Interval Lower Bound	Confidence Interval Upper Bound
Intercept	6858.4	4365.0	9351.8
Year	-6899.6	-10592.4	-3206.7
Banking/Banks	-6868.3	-10464.3	-3272.2
Broadcasting (Television, Radio, Internet)	-6545.3	-10288.0	-2802.6
Communications	-6444.2	-10560.6	-2327.7
Crown Corporations	1498.1	-1956.8	4953.0
Energy/Mining/Mineral Processing	-3481.4	-24732.5	17769.7
Federal Public Services/Public Service Departments/Crown Corporations	-4299.8	-7708.2	-891.3
Feed, Flour and Seed	-5831.6	-9165.8	-2497.4
Grain Handling/Grain Elevators	-6384.7	-9793.8	-2975.6
Indigenous	-6479.7	-10127.4	-2831.9
Interprovincial Infrastructure (Bridges, Tunnels, Canals, Causeways)	12705.9	9084.2	16327.5
Long shoring/Stevedoring/Port/Harbour Operations/Pilotage	-6303.4	-9723.9	-2882.9
Pipeline Transportation	-512.9	-4054.7	3028.9
Postal Services/Postal Contractors	-6583.9	-12730.4	-437.4
Public Service Departments	-3174.3	-6579.9	231.3
Rail Transportation	-4923.1	-8358.9	-1487.3
Road Transportation	1559.5	-2044.5	5163.6
Water Transportation (Shipping and Ferries)	-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

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

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-test?**), 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-test?**), Model 2 has slightly lower AIC and BIC than Model 1, implying that Model 2 has more prediction power.

Table 2: 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 × Year		3.40
		(0.93)
IndustryBroadcasting (Television, Radio, Internet) × Year		3.39
		(0.90)
IndustryCommunications × Year		3.23
		(0.94)
IndustryCrown Corporations × Year		3.18
		(1.03)
IndustryEnergy/Mining/Mineral Processing × Year		-0.76
		(0.87)
IndustryFederal Public Services/Public Service Departments/Crown Corporations × Year		1.71
		(5.32)
IndustryFeed, Flour and Seed × Year		2.13
		(0.86)
IndustryGrain Handling/Grain Elevators × Year		2.88
		(0.84)
IndustryIndigenous × Year		3.15
		(0.86)

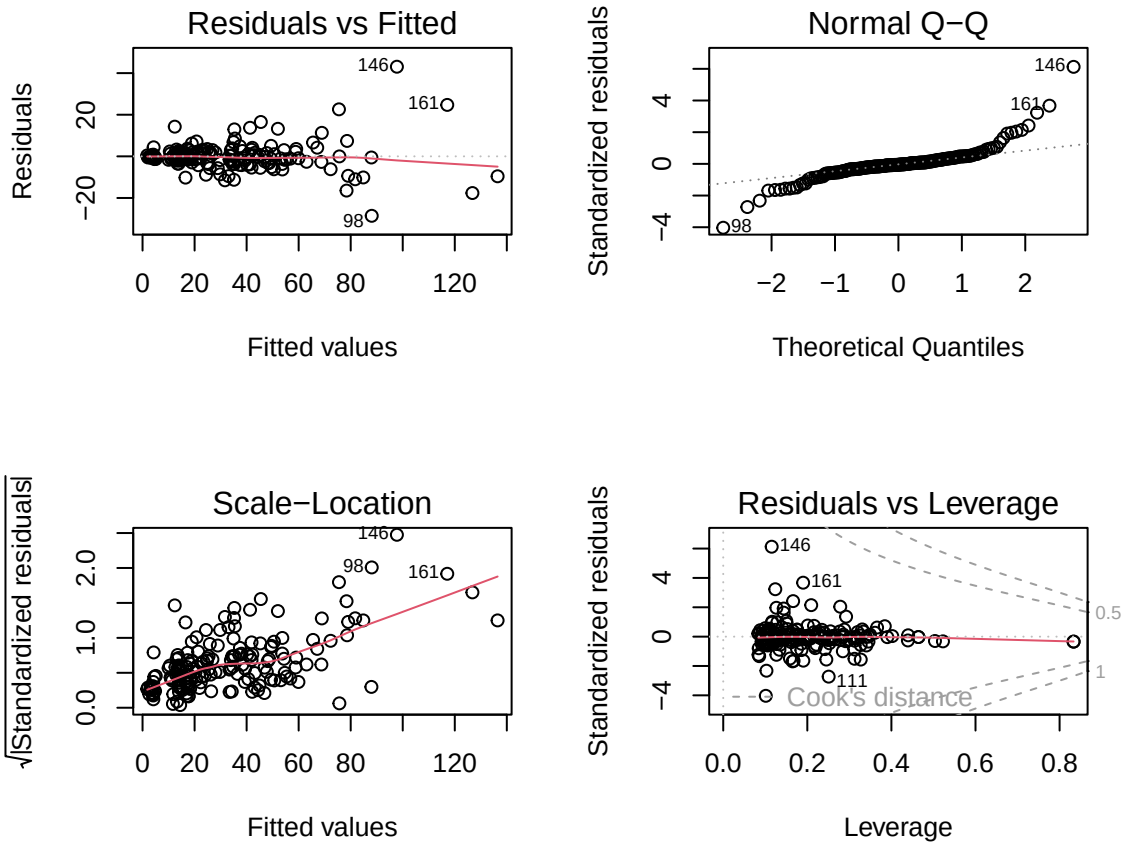
	Model 1	Model 2
IndustryInterprovincial Infrastructure (Bridges, Tunnels, Canals, Causeways) × Year		3.21
		(0.92)
IndustryLong shoring/Stevedoring/Port/Harbour Operations/Pilotage × Year		-6.30
		(0.91)
IndustryPipeline Transportation × Year		3.10
		(0.86)
IndustryPostal Services/Postal Contractors × 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.	-663.324	-576.767
RMSE	10.95	6.66

Table 3: 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 (**table3?**), 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



Check for assumptions: Assumptions to the multiple linear regression model are checked to ensure that the model is valid for this dataset. (**figure3?**) 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.

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.

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