

---

# **DSCI 310: Historical Horse Population in Canada**

**Tiffany Timbers & Jordan Bourak**

**Mar 18, 2022**



## CONTENTS



This project explores the historical population of horses in Canada between 1906 and 1972 for each province.

## Data

Horse population data were sourced from the [Government of Canada's Open Data website](#). Specifically, (Government of Canada, 2017)<sup>1</sup> and (Government of Canada, 2017)<sup>2</sup>.

## Methods

The R programming language (R Core Team, 2019) and the following R packages were used to perform the analysis: knitr (Xie, 2014), tidyverse (Wickham, 2017), and bookdown (Xie, 2016) *Note: this report is adapted from (Timbers, 2020).*

## Results

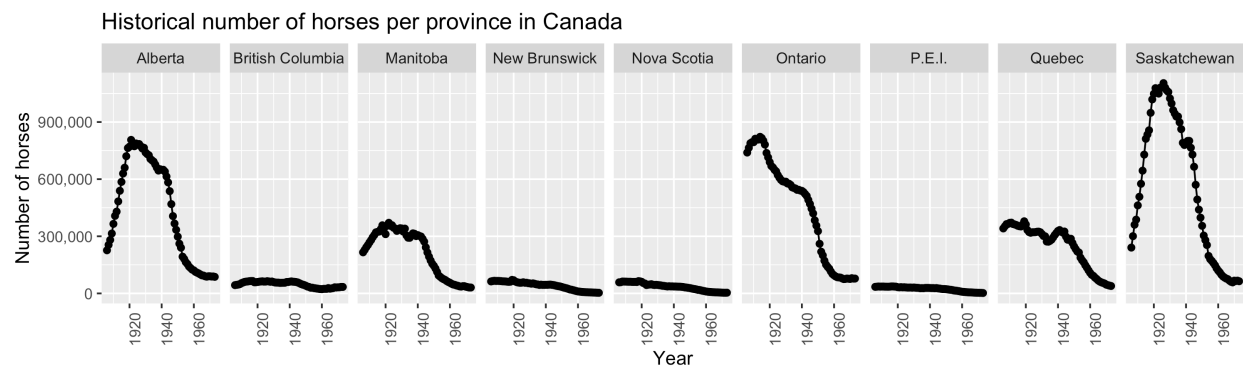


Fig. 1 Horse populations for all provinces in Canada from 1906 - 1972

We can see from Fig. 1 that Ontario, Saskatchewan and Alberta have had the highest horse populations in Canada. All provinces have had a decline in horse populations since 1940. This is likely due to the rebound of the Canadian automotive industry after the Great Depression and the Second World War. An interesting follow-up visualisation would be car sales per year for each Province over the time period visualised above to further support this hypothesis.

5pt

```
horses_sd = pd.read_csv("../results/horses_sd.csv")

largest_sd_prov = str(horses_sd['Province'][0])
glue("largest-sd-prov", largest_sd_prov)

horses_sd_noindex = horses_sd.style.hide_index()
glue("horses-tbl", horses_sd_noindex)
```

20pt

'Saskatchewan'

```
<pandas.io.formats.style.Styler at 0x105a00df0>
```

Suppose we were interested in looking in more closely at the province with the highest spread (in terms of standard deviation) of horse populations. We present the standard deviations here:

20pt

```
<pandas.io.formats.style.Styler at 0x105a00df0>
```

Table 1: Standard deviation of number of horses for each province between 1940 - 1972

Note that we define standard deviation (of a sample) as:

$$s = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 / (n - 1)}.$$

Additionally, note that in Fig 2. we consider the sample standard deviation of the number of horses during the same time span as Fig. 1.

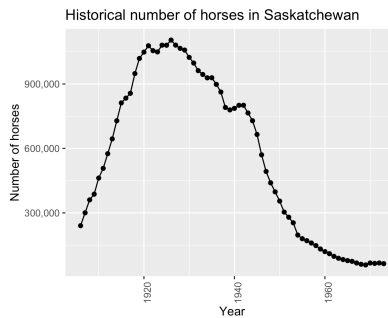


Fig. 1: Horse populations for the province with the largest standard deviation

In Fig. {number} we zoom in on the province of ?, which had the largest spread of values in terms of standard deviation.