Results

Tiffany Timbers & Jordan Bourak

Contents

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   1.0.1
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

This is citation, R Core Team (2019)

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

Horse population data were sourced from the Government of Canada's Open Data website (Government of Canada, 2017a and Government of Canada, 2017b).

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).*

```
knitr::include_graphics("../../results/horse_pops_plot.png")
```
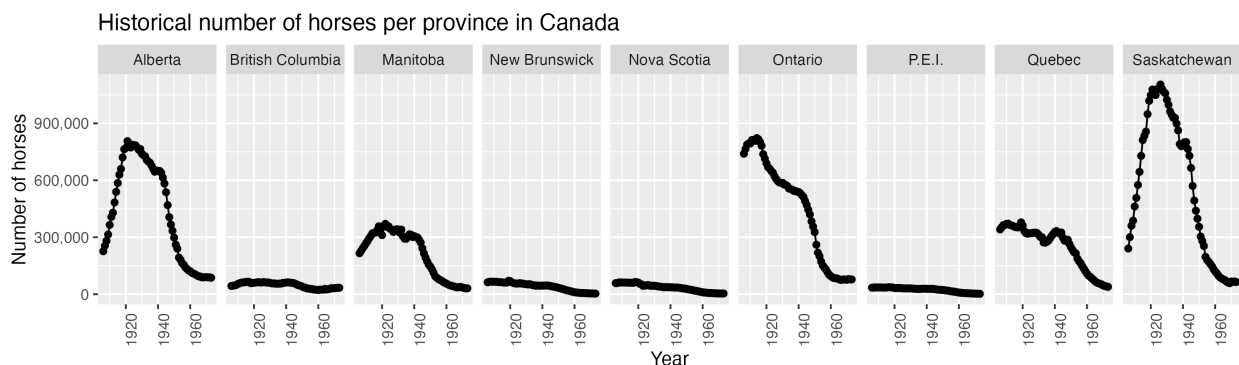

Figure 1: Horse populations for all provinces in Canada from 1906 - 1972

We can see from Figure 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.

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:

```
horses_sd_table <- read_csv("../../results/horses_sd.csv")
```

```
## Rows: 9 Columns: 2
## -- Column specification ----------------------------------------------------------
## Delimiter: ","
## chr (1): Province
## dbl (1): Std
##
## 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.
```

```
largest_sd <- horses_sd_table$Province[1]
knitr::kable(horses_sd_table)
```

Province	Std
Saskatchewan	377265.58
Ontario	266435.32
Alberta	266063.19
Manitoba	122403.87
Quebec	111411.10
New Brunswick	22019.49
Nova Scotia	19879.25
British Columbia	14945.66
P.E.I.	11355.75

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

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

Additionally, note that in Table 1 we consider the sample standard deviation of the number of horses during the same time span as Figure1.

```
knitr::include_graphics("../../results/horse_pop_plot_largest_sd.png")
```
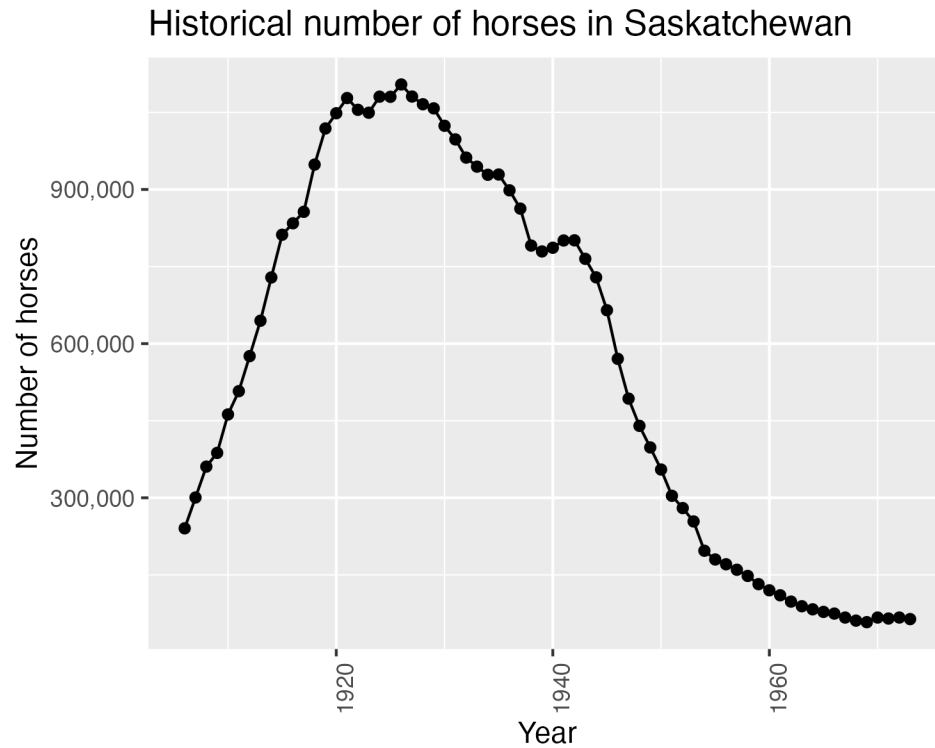
Figure 2: Horse populations for the province with the largest standard deviation

In Figure 2 we zoom in and look at the province of ?, which had the largest spread of values in terms of standard deviation.

References

R Core Team. 2019. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.