Houghton Snowfall Data Analysis

By: Belee Pandya

Language & Packages

- **Language used:** R Programming
- **Software:** R Studio
- **Packages Used:** ggplot2, datatable, forecast, neuralnet

History

- It is the 59th snowiest city in the United States.
- Houghton has had an average annual snowfall of 207.7 inches over the last 30 years
- 73% of Houghton's snow fell during the winter months

Source: www.mtu.edu

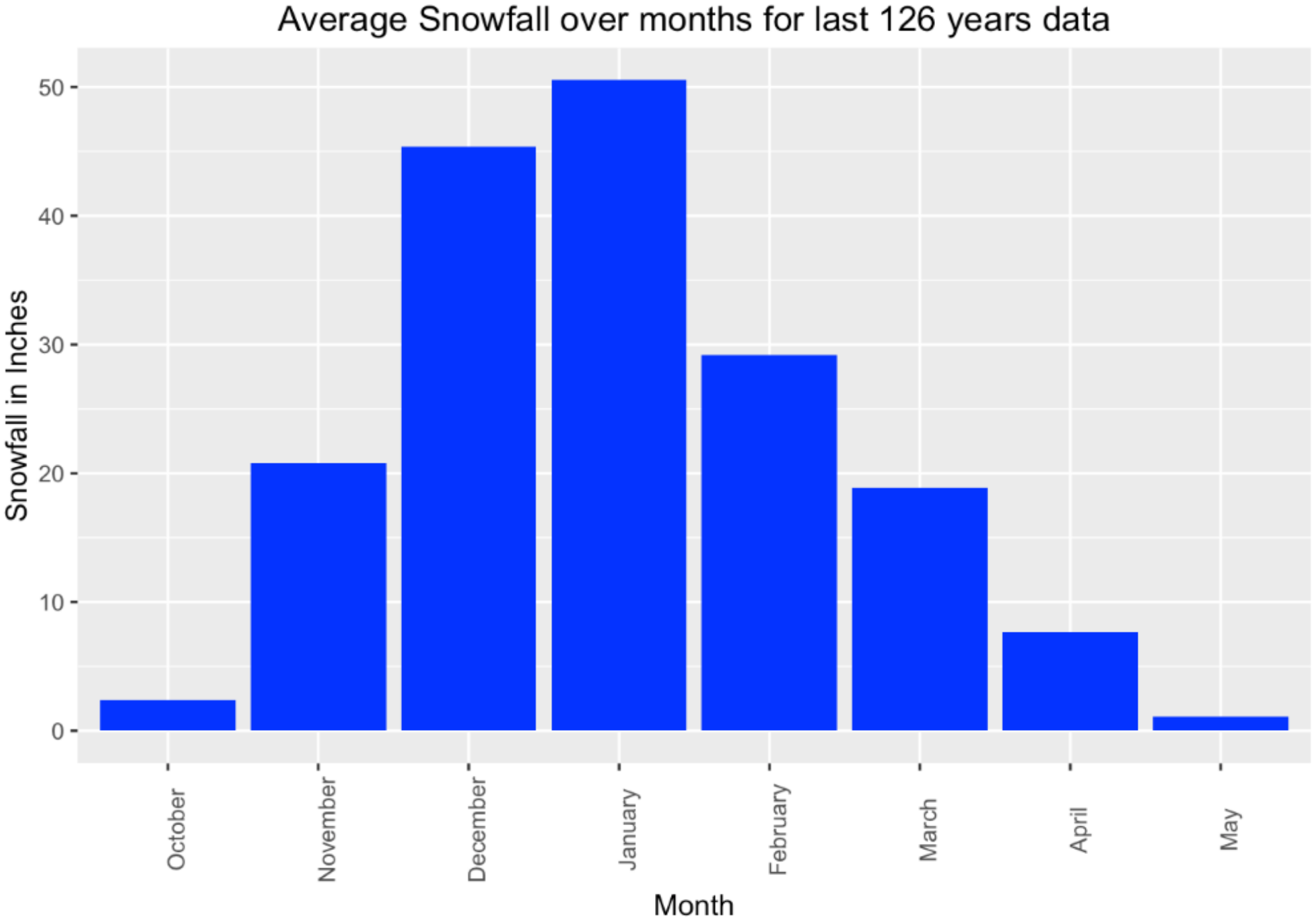
Goal

Maximum and Minimum snowfall till today.

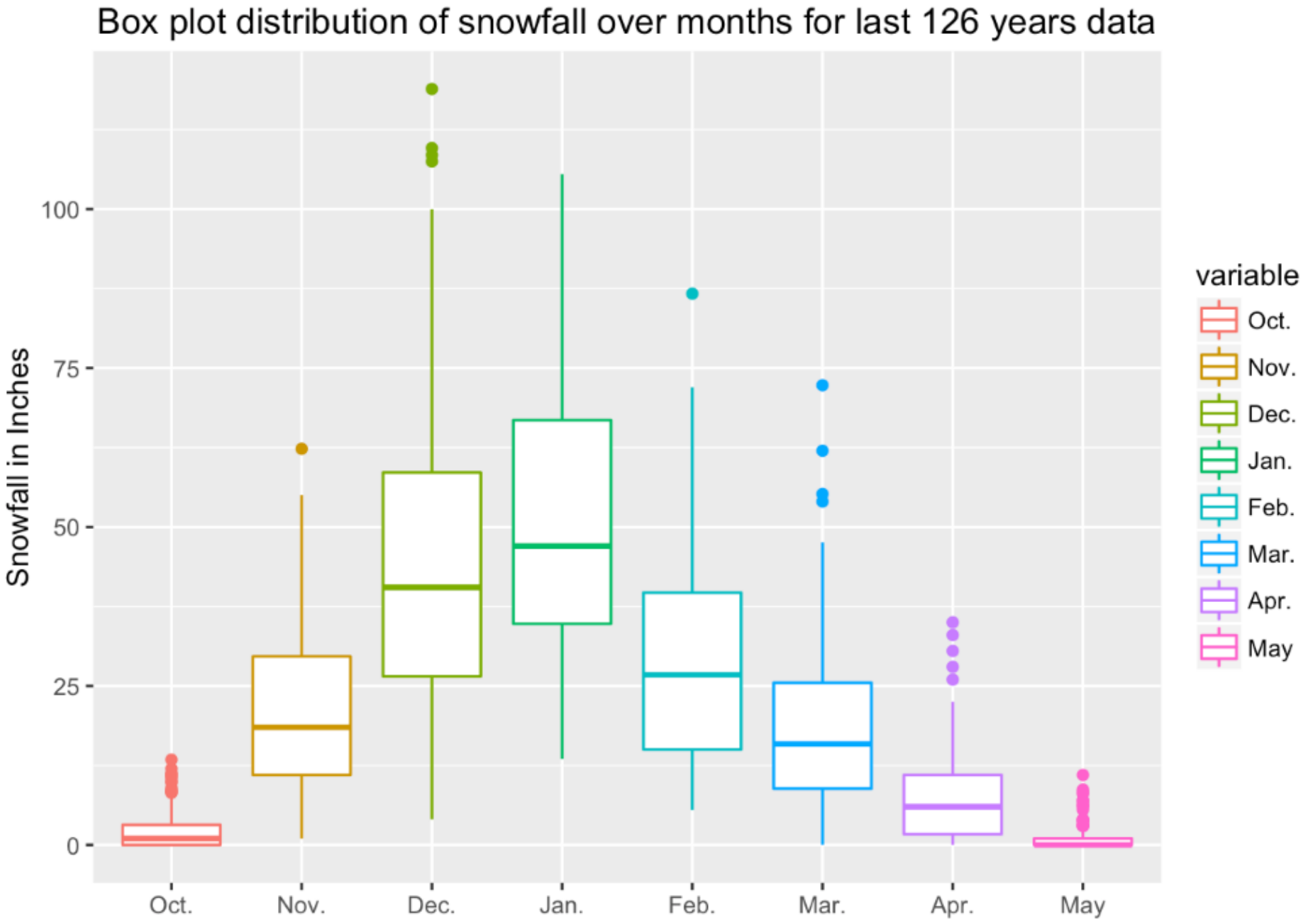
Prediction of snowfall for current and upcoming years.

Cross validation of various models considering current dataset.

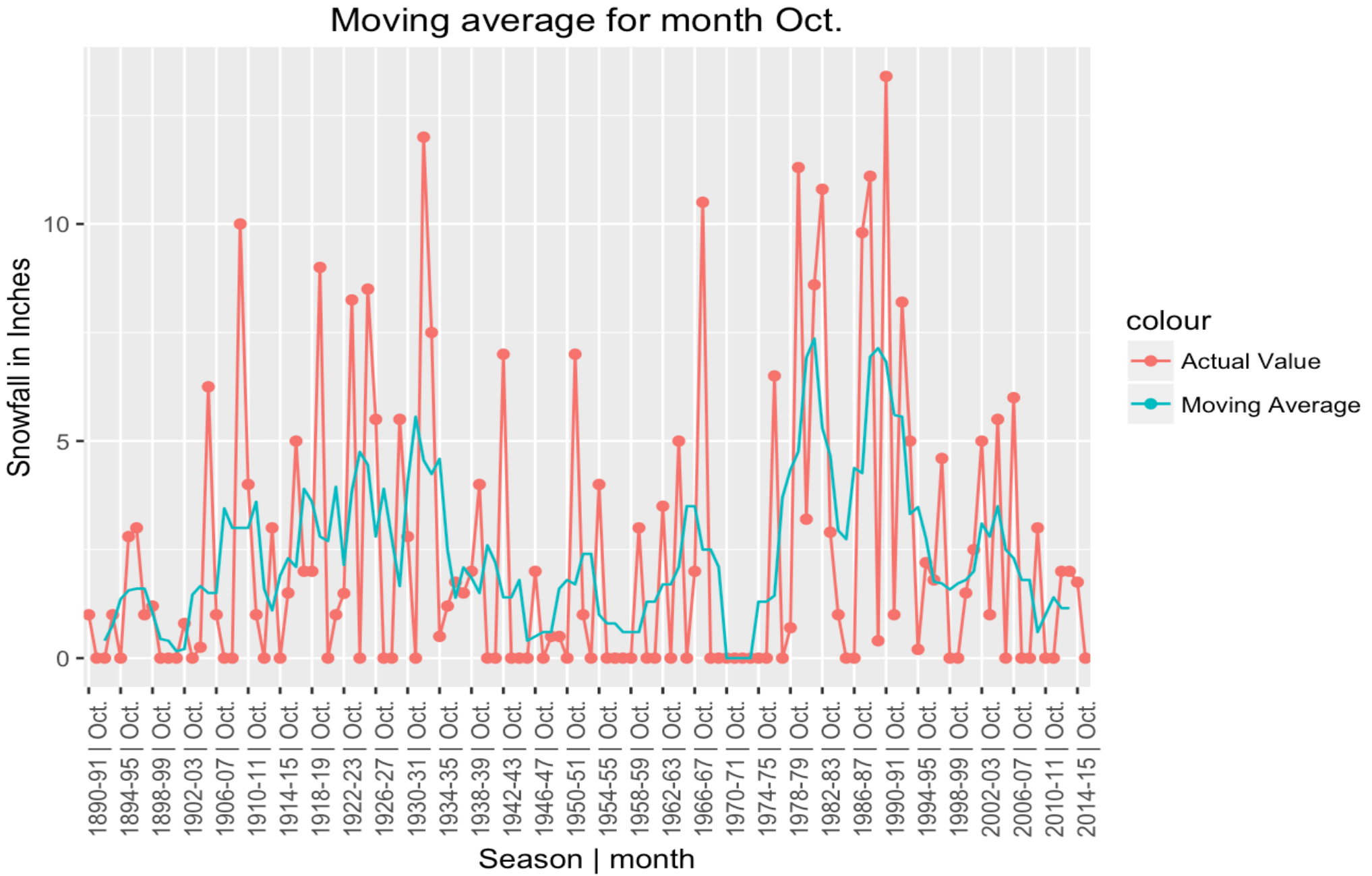
Average snowfall over months



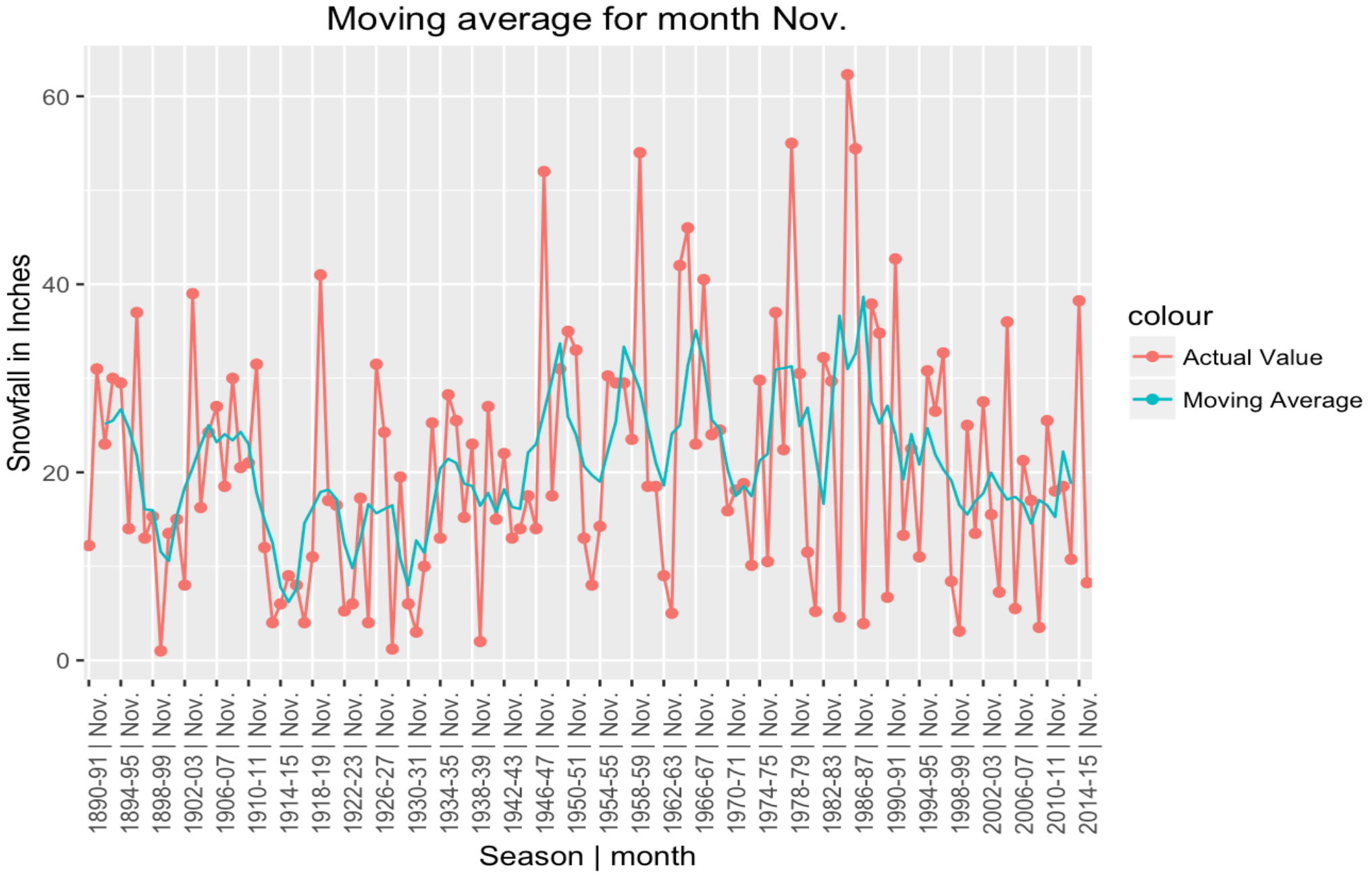
Distribution of snowfall over months



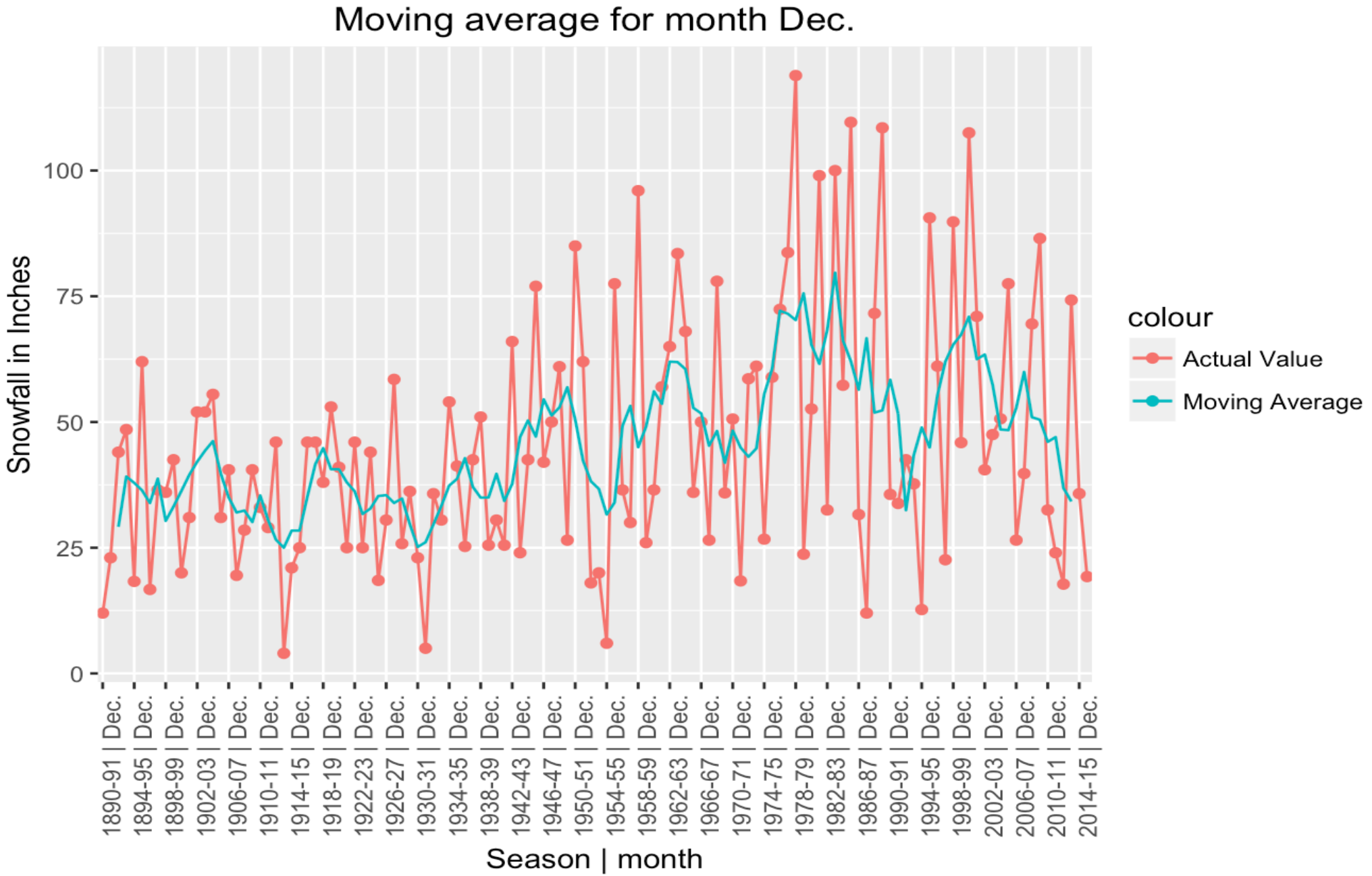
Moving average for October



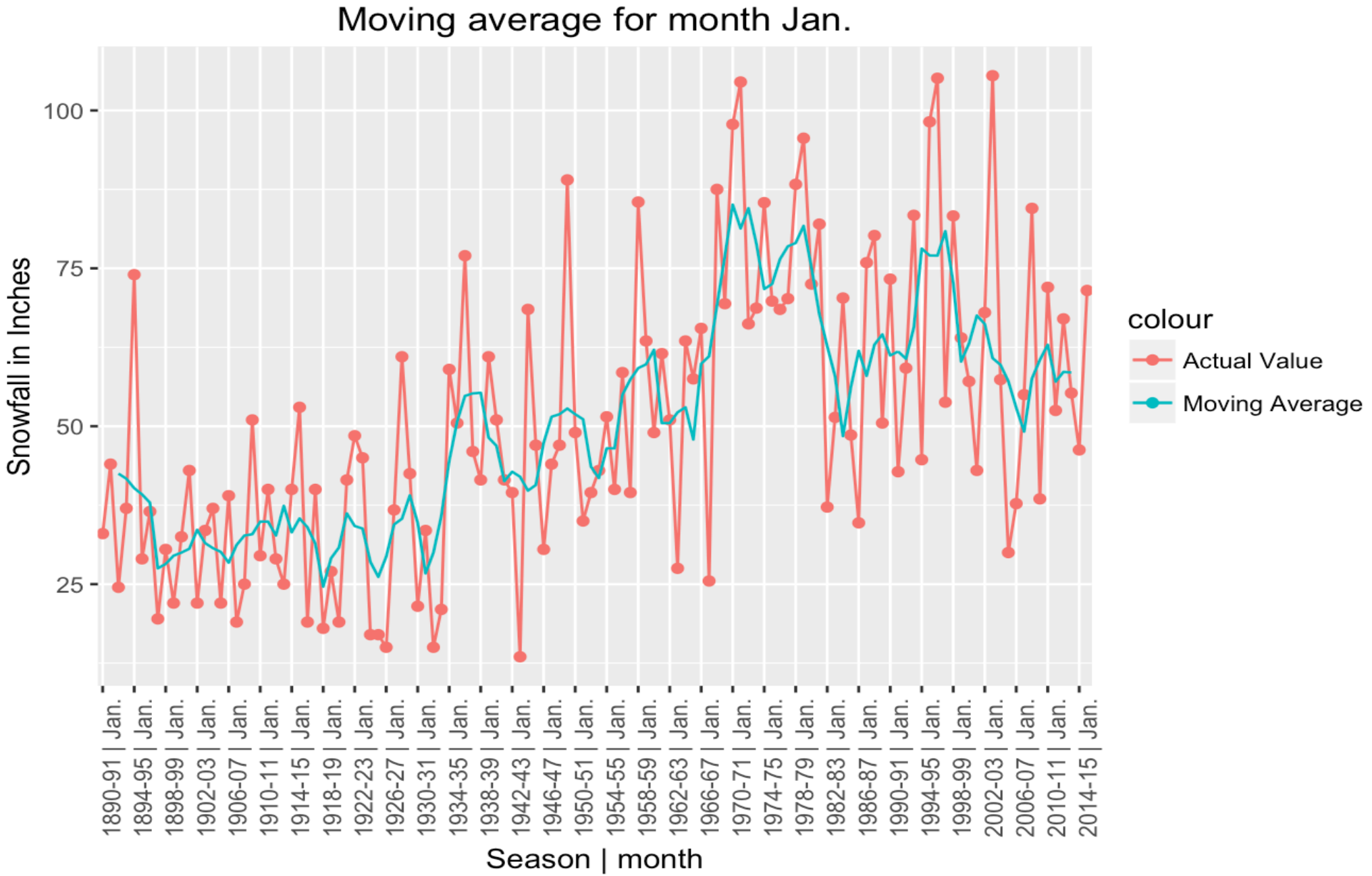
Moving average for November



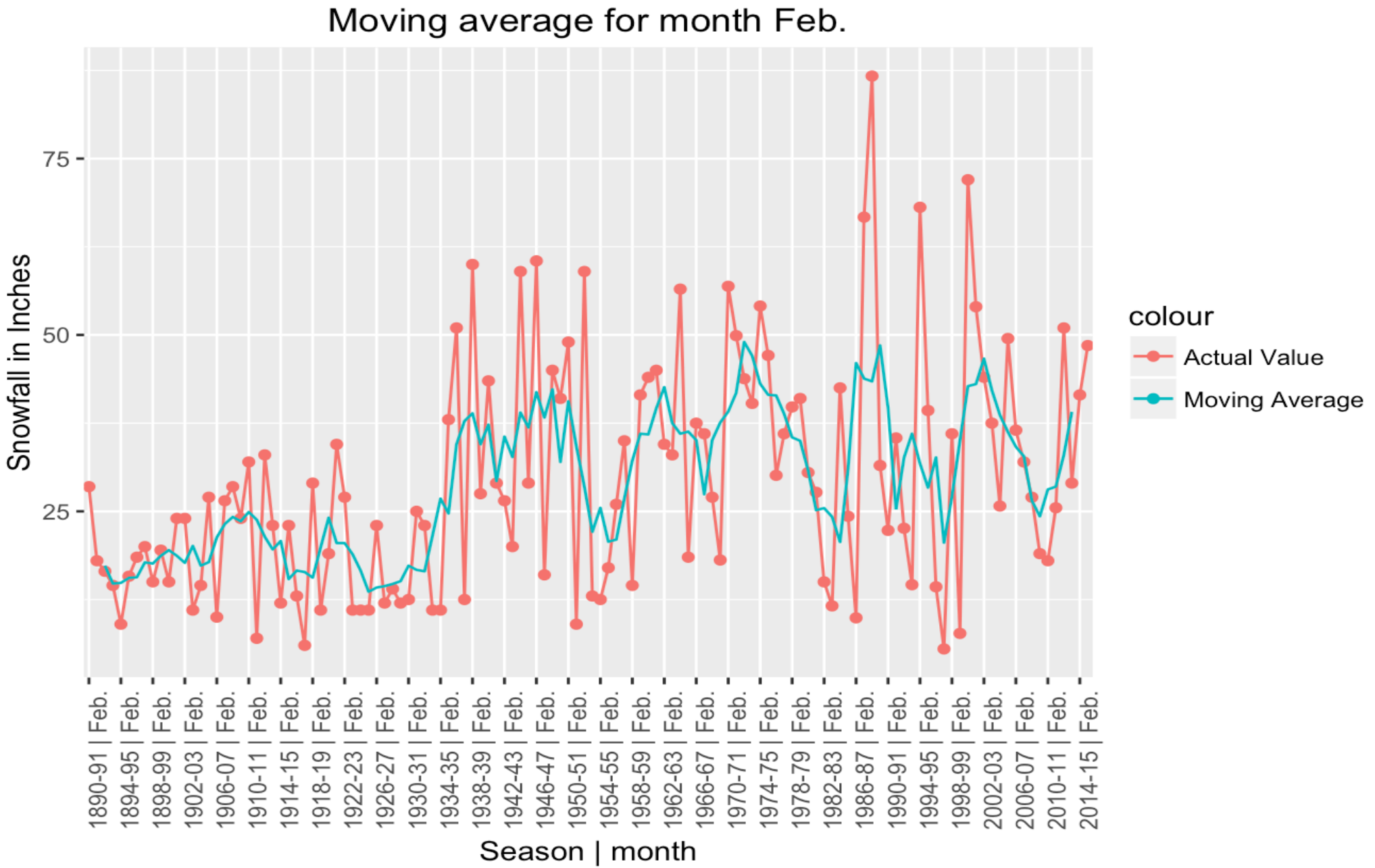
Moving average for December



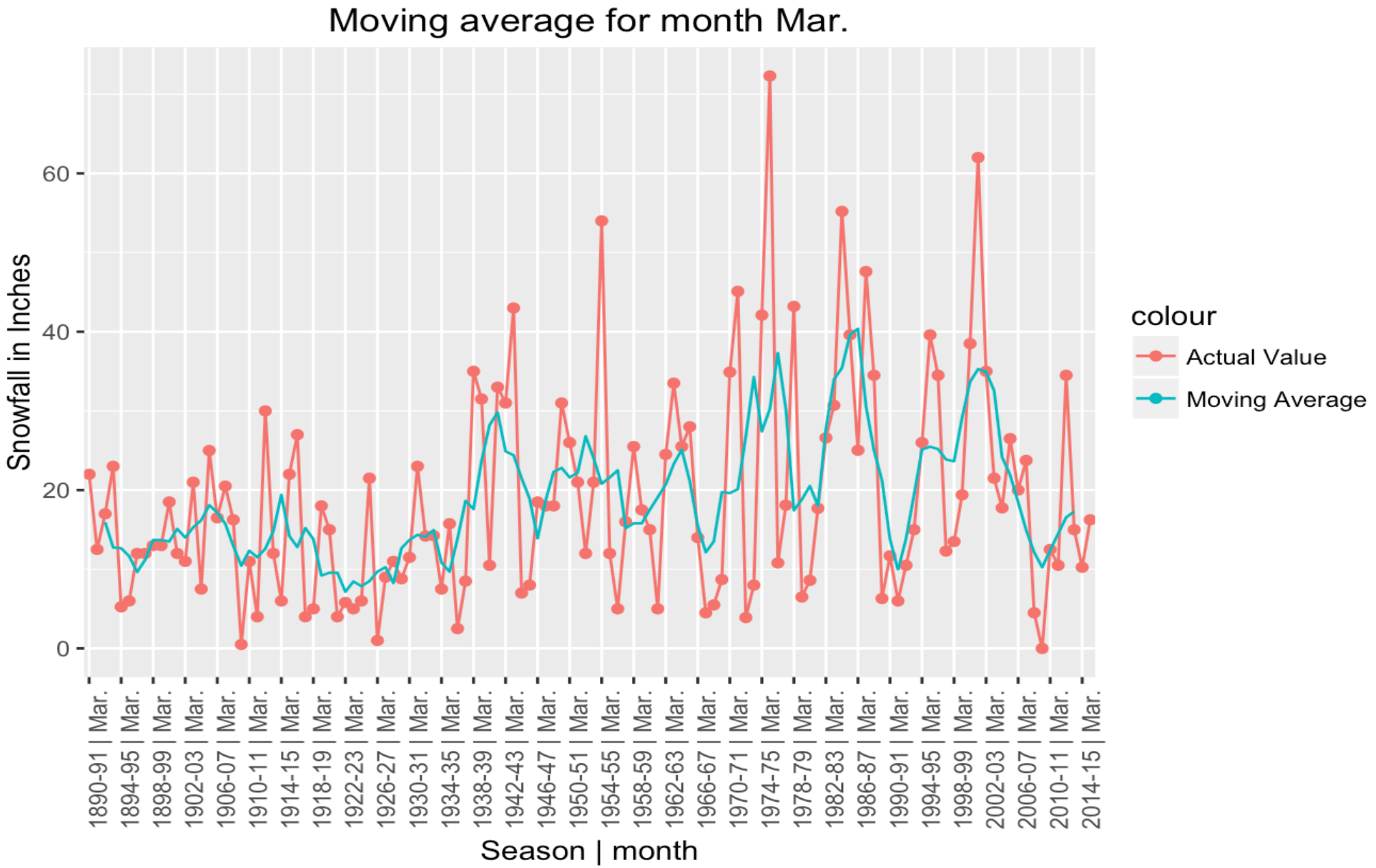
Moving Average of January



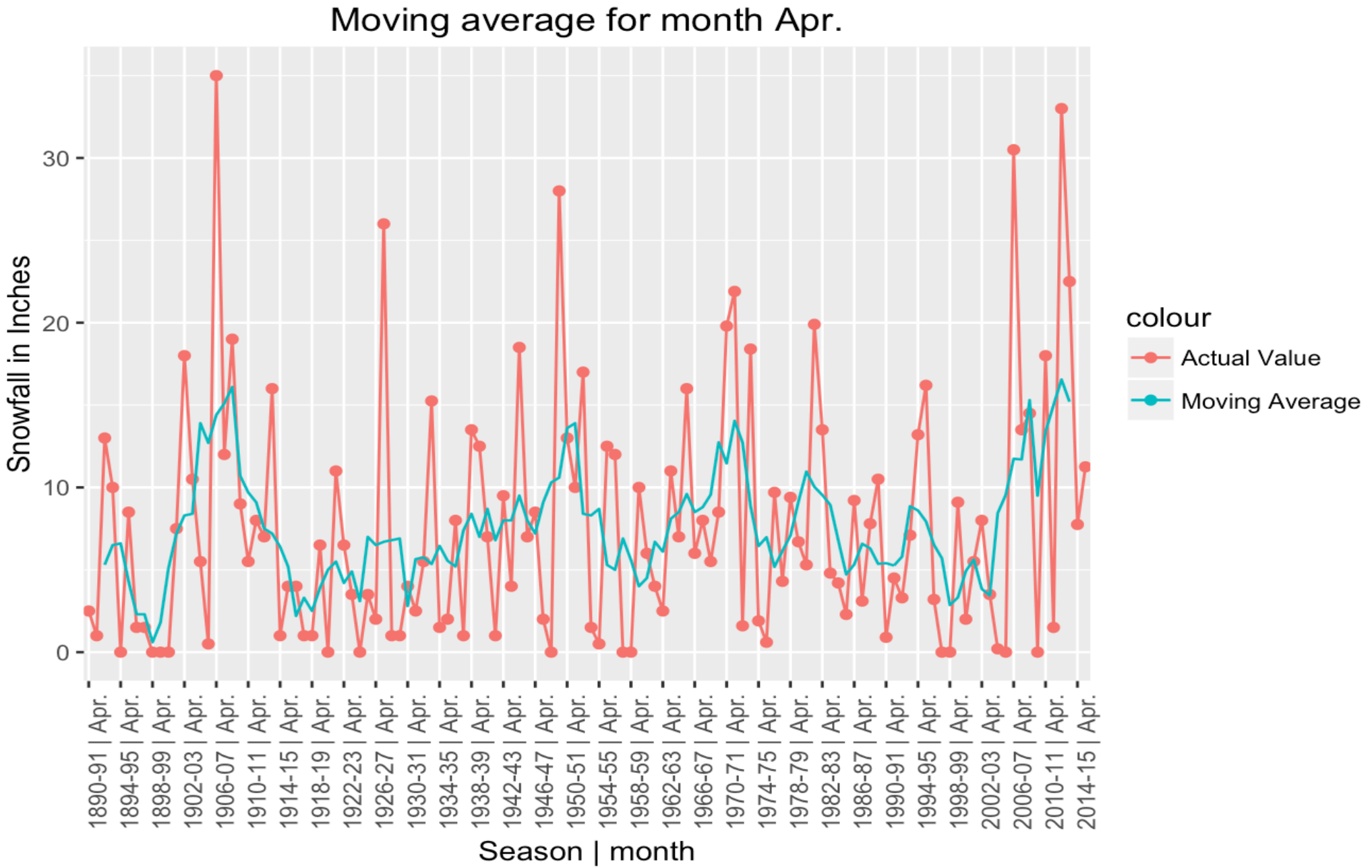
Moving Average for February



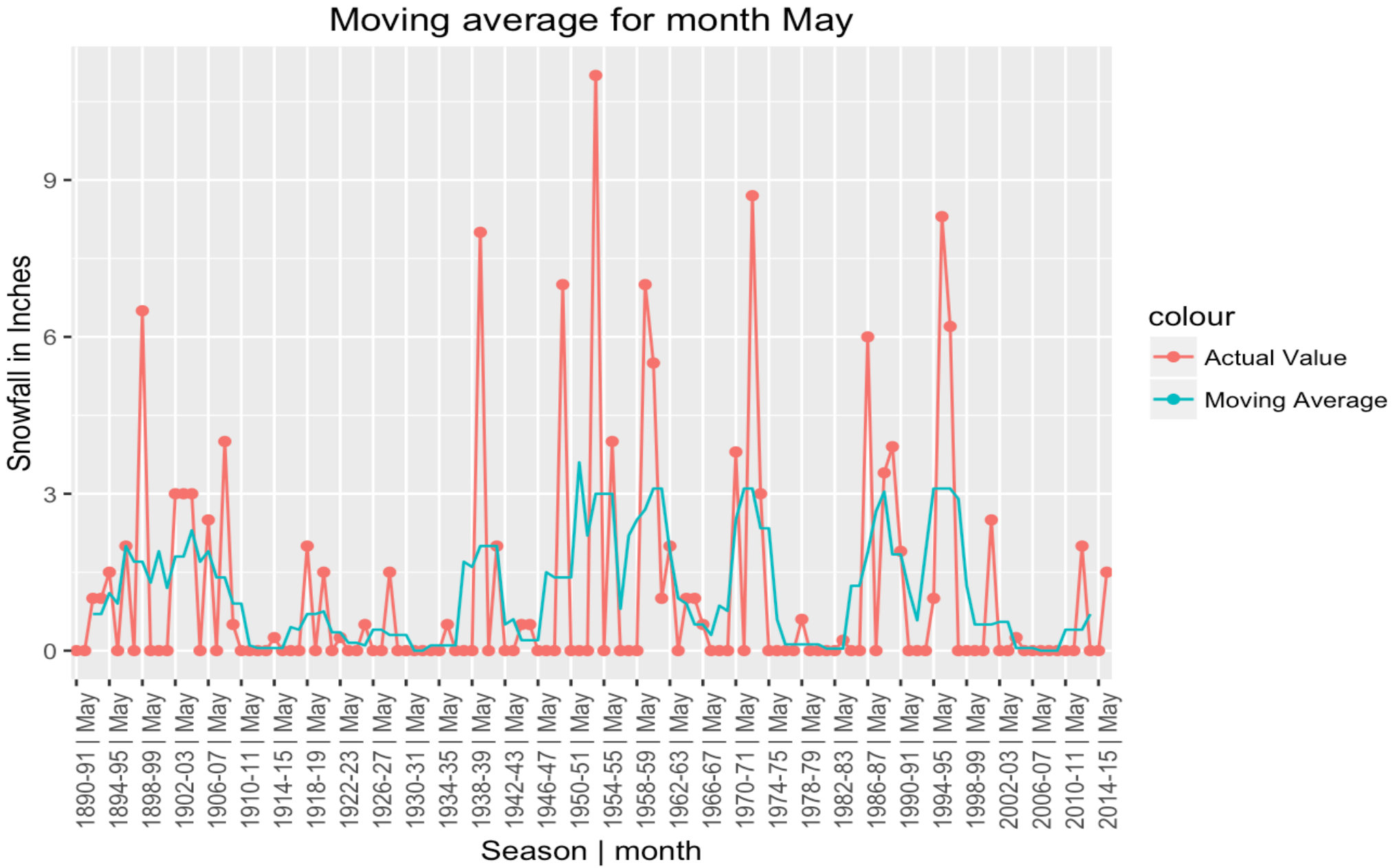
Moving average for March



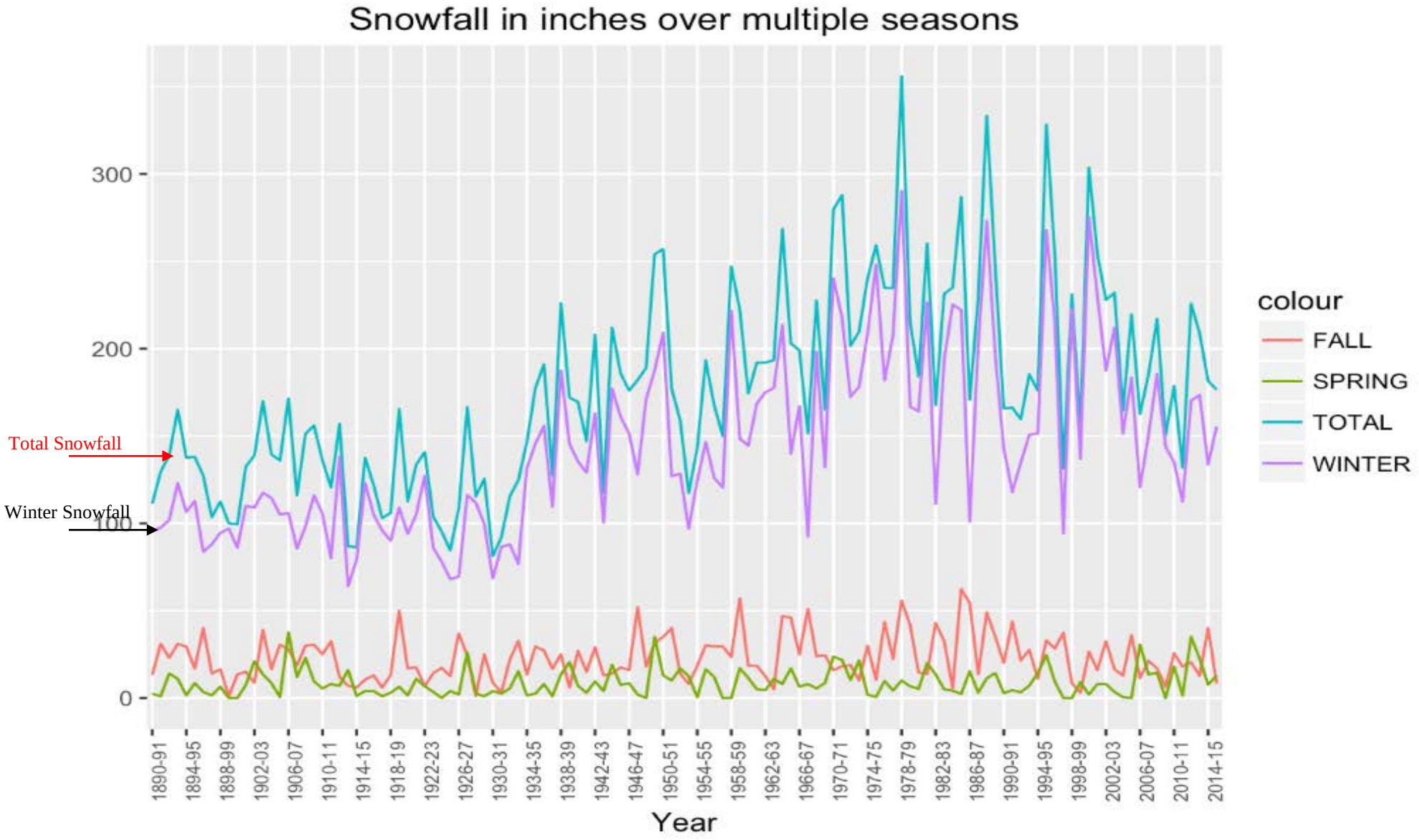
Moving average for April



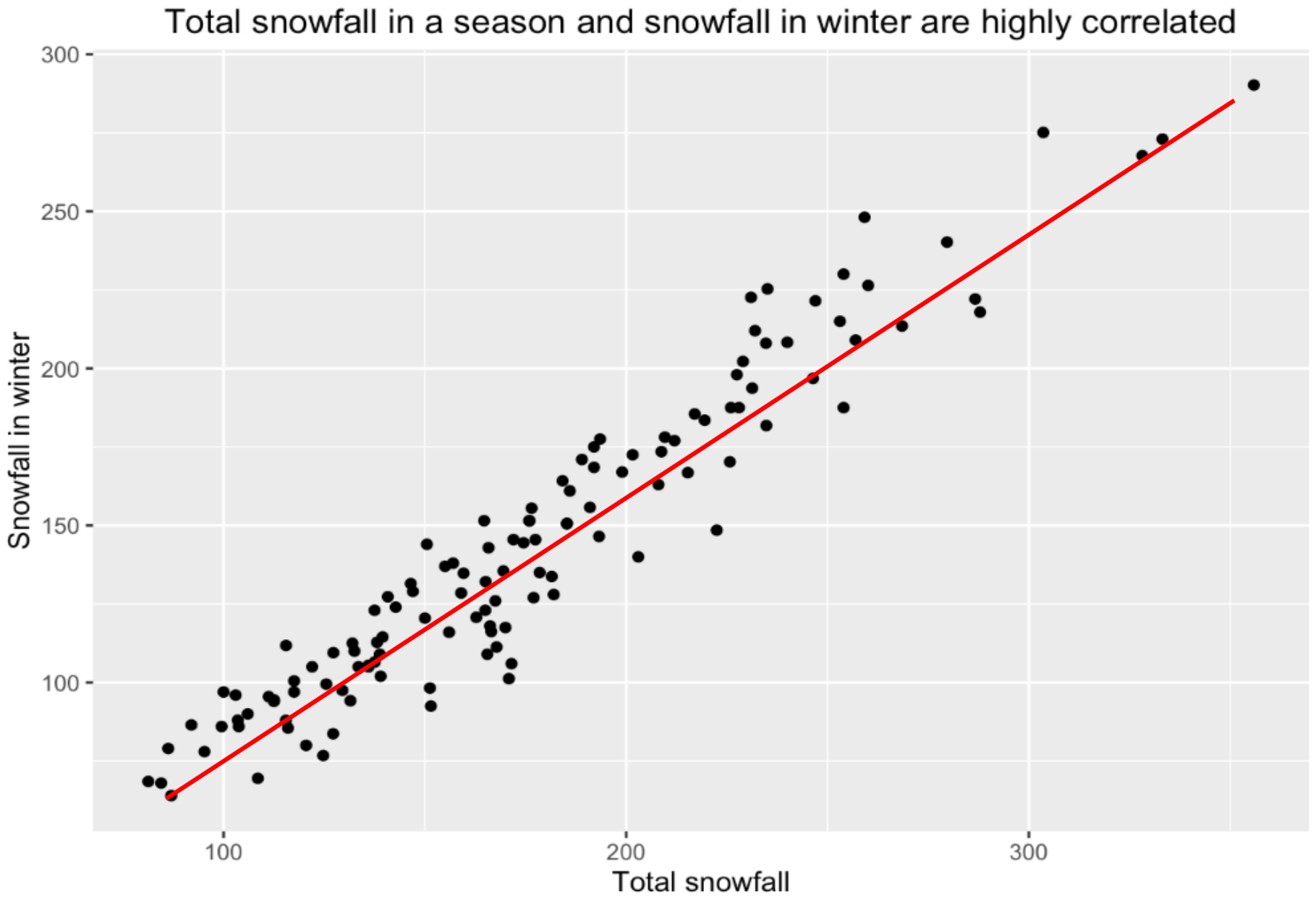
Moving average for May



Correlation between total snowfall and snowfall in winter (December - March)



Correlation between total snowfall and snowfall in winter (December - March)



Prediction of snowfall for the current and upcoming seasons

Multiple techniques can be used to predict or forecast the snowfall in future. These techniques can be grouped into two categories :

Statistical Models

- Moving Average (over last n years data)
- Autoregressive Moving Average Model (over last n years data)

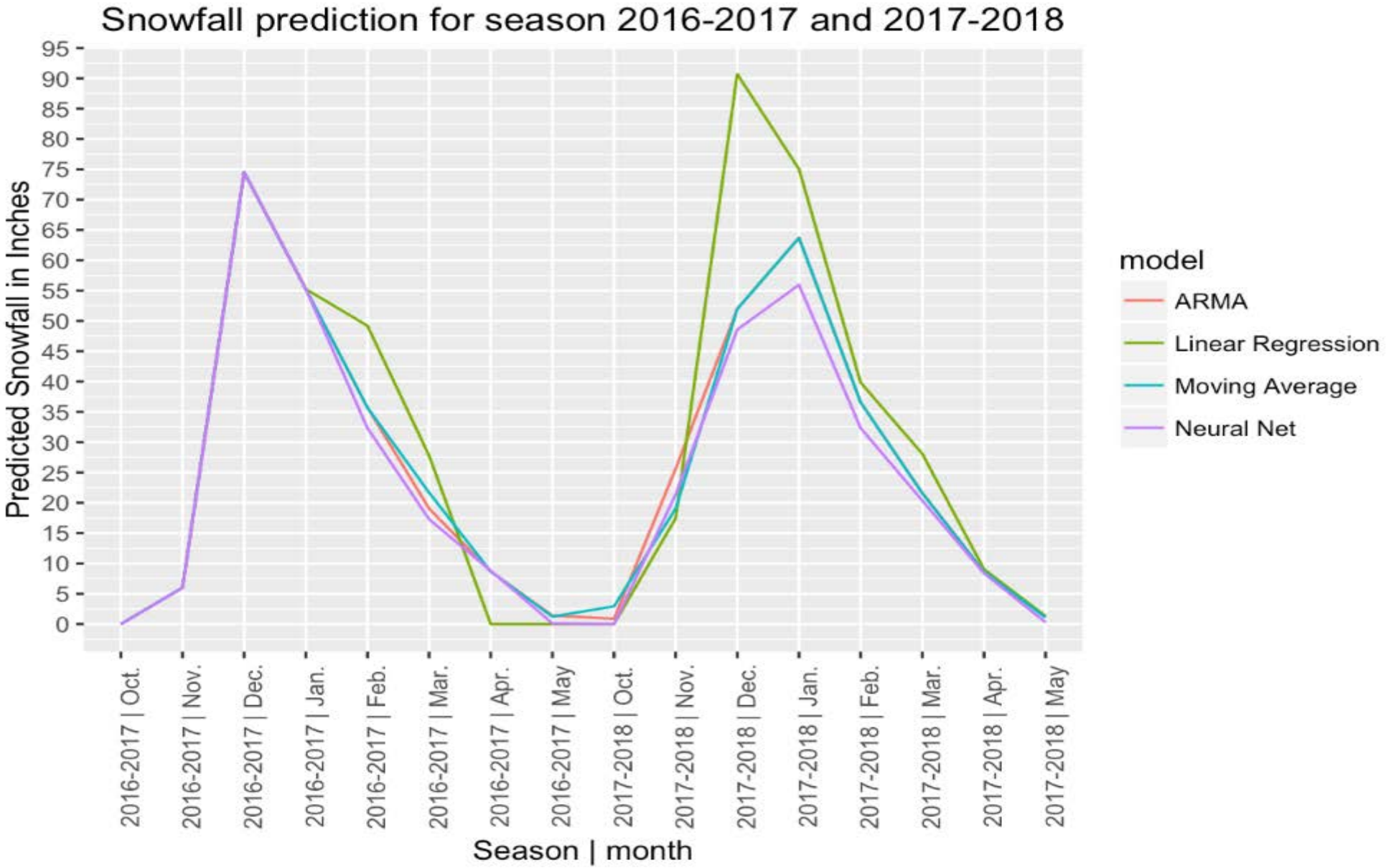
and so on…

Predictive learning / Regression Models

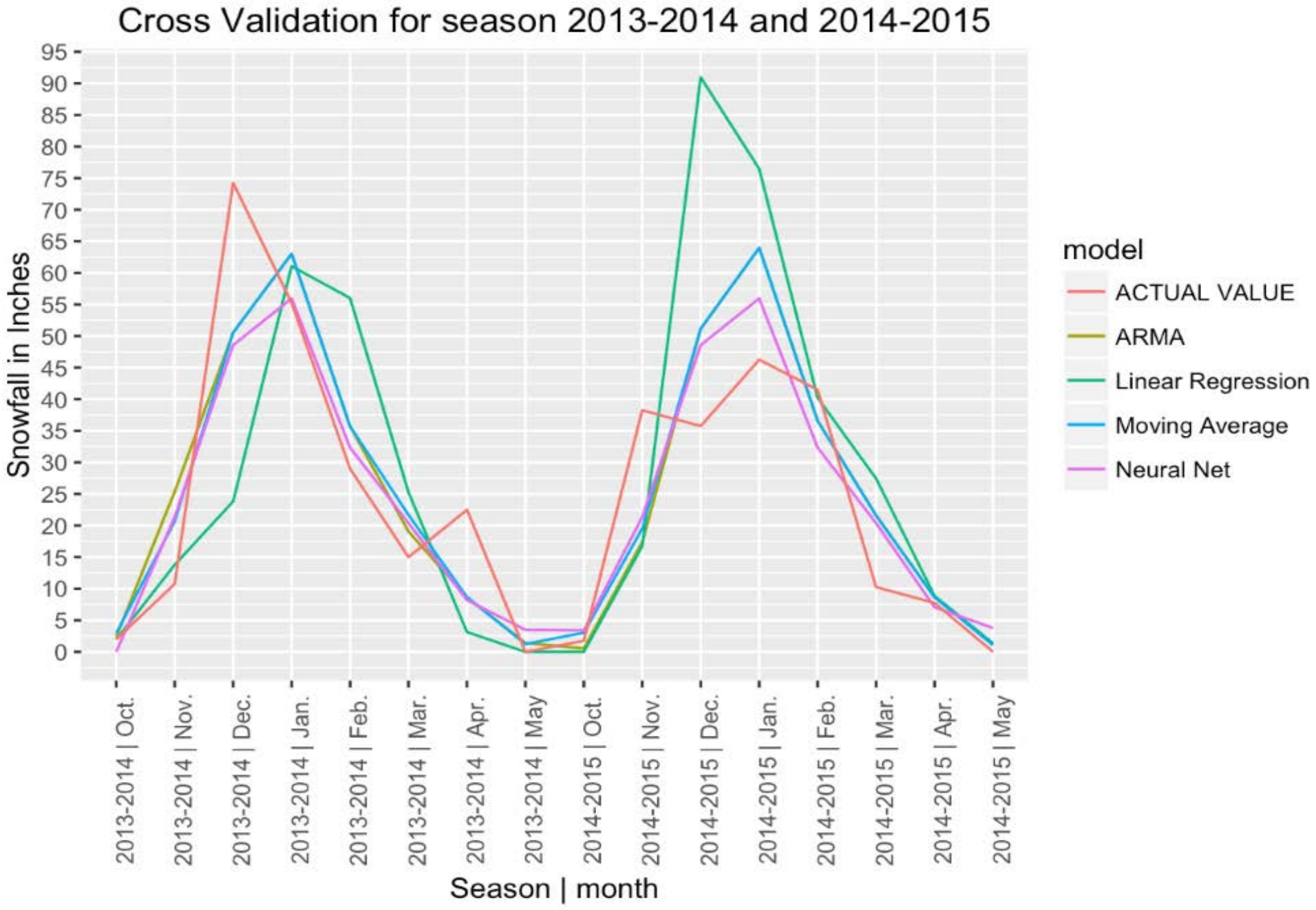
- Linear Regression (over last n years data and this)
- Neural Nets

and so on…

Prediction of snowfall for the current and upcoming seasons



Cross Validation performance of various models considered above

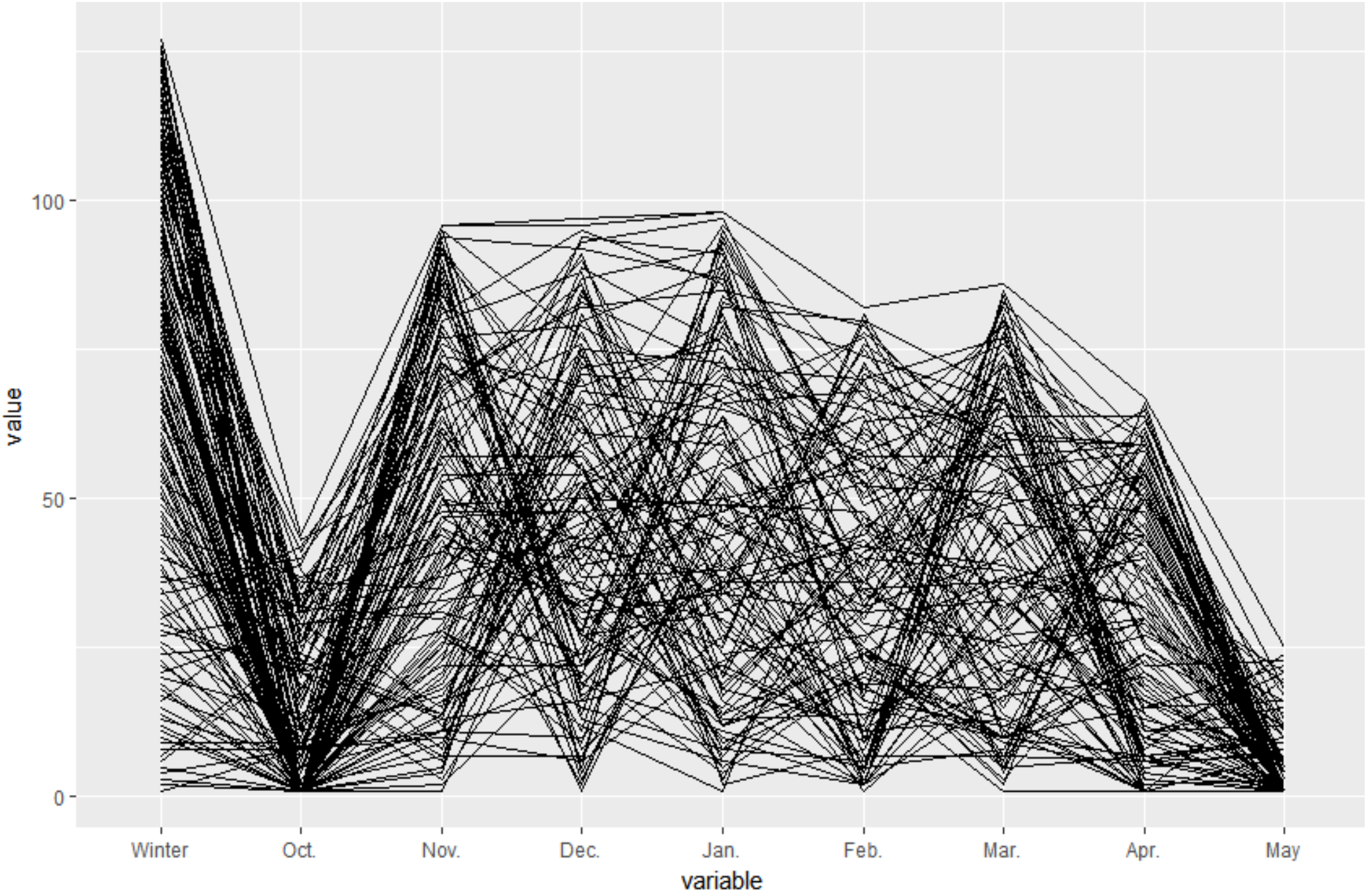


Drawbacks of prediction models

1. Accuracy over prediction decreases, i.e., for a fixed margin of error, our confidence score for a particular prediction of snowfall decreases when we go further away from current time

2. As we predict far away in future, our training data will start containing instances of our own predictions, which will lead to erroneous predictions and degrade model accuracy over time.

Parallel Coordinates



Summary

Characteristics:

- January has maximum snowfall
- May has minimum snowfall
- Average snowfall is 19.5 Inches

Correlation:

- Correlation coefficient is 0.96 between total snowfall in a season and snowfall during winter for last 126 years data.
- Correlation coefficient is 0.36 when considered for fall or spring.

Predictions:

- Highest snowfall in December – 75’’ (2016-2017)
- Highest snowfall in December – 92’’ (2017-2018)

Thank you!