STAT394 Group 1 Final Report: A comparison of seasonal classifications for Wellington and Auckland weather

Jordan Monk, Andrea Knox

16 October, 2022

Contents

1 Abstract

Wellington's weather is notoriously unpredictable and in 2014 Adam Shand proposed that Wellington's weather would be more realistically described by a new and unconventional seasonal classification that splits spring into two periods (August and December) and renames the period from September to November as "Shitsville" (Shand, A 2014). This proposal has received support, with several hundred retweets and a website based upon it (Shand, A and Ivshin, D 2022). We investigated whether these "real" seasons in fact provide a better classification of the weather using daily Wellington and Auckland weather data from 2017 to 2021. If the real seasons performed better, it would suggest that serious consideration should be given to a proposal to change to the seasonal naming system.

We used four different methods and across all four we obtained similar results. Silhouette plots suggested that the conventional seasons are better than real seasons at grouping together days with similar weather. We found larger distances between seasonal summary statistics and lower standard deviations for conventional seasons than real seasons, suggesting that conventional seasons are better at separating similar observations into more separate groups. And finally we found that k-means clustering based on daily weather data groups days into categories that are more similar to conventional seasons than real seasons. These results held true for Wellington and Auckland and we conclude that our evidence does not support a proposal to change the seasonal naming system.

We also found that that neither seasonal classification performed particularly well at partitioning days into groups with similar weather. This concords with what is likely be most New Zealanders' subjective experience of unpredictable weather, no matter what the season. Our findings remind us to acknowledge that the seasons are in fact based on the tilt of the Earth's axis relative to the sun, with weather an incidental outcome of this tilt rather than a key driver of how seasons are classified.

2 Introduction

Wellington is notorious for its windy and unpredictable weather. The city has large public artworks celebrating the wind (Figure 1) and several classic New Zealand popular songs refer to Wellington's weather with lyrics such as, *"I wish I was in Wellington, the weather's not so great."* (The Mutton Birds 1994) and *"She said the weather's crap, I said you'll come around … You can't beat Wellington on a good day."* (The Datsun Violets 1996).

Figure 1: The 'blown away' sign in Miramar, Wellington. Image by Wainuiomartian, CC BY-SA 4.0 https://creativecommons.org/licenses/by-sa/4.0, via Wikimedia Commons

Anecdotally, the weather in Wellington does not appear to be clearly demarcated by the seasons and in 2014 Adam Shand posted a tweet proposing an alternative way of classifying Wellington's seasons (Shand, A 2014). We reproduce this alternative classification in Figure 2, alongside the conventional seasons for comparison. Adam Shand's 2014 tweet has been retweeted several hundred times and has given rise to a website that displays the "realistic" season for the current date (Shand, A and Ivshin, D 2022). Interestingly, the website refers to the schematic as *"A Realistic New Zealand Calendar"*, suggesting that the authors now hold the view that it applies to all of New Zealand, not just Wellington.

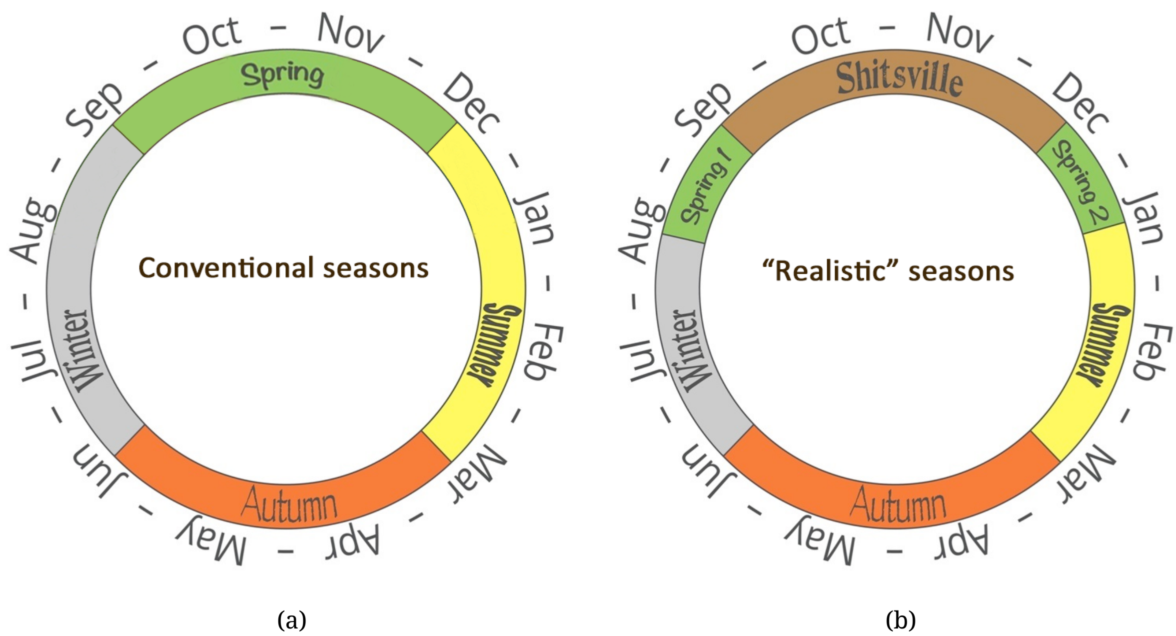


Figure 2: Schematic of the conventional seasons (a), and the alternative classification proposed by Adam Shand (b). Images are adapted from https://www.realnzweather.com.

In this report we use five years of weather data from Wellington and Auckland cities and we investigate, for both cities, which seasonal classification better fits the weather data: the conventional seasons, or the realistic seasons proposed by Shand, A (2014). We ask the question: *should we change the seasonal naming system for Wellington and/or Auckland?*

This gives rise to the two key sub-questions that the analysis in this report focuses on:

- Which classification better partitions observed weather patterns in Wellington?
- Which classification better partitions observed weather patterns in Auckland?

These questions have important public policy implications. If the realistic seasons are a better classifier it would establish the beginnings of an evidence base supporting a change to the seasonal naming system for New Zealand or for specific cities. While such a change could be disruptive, it could also capture public imagination in a similar manner to the Laser Kiwi flag (Ministry for Culture and Heritage 2021).

In this report we investigate the relative performance of each classification at grouping together days with similar weather values. We assess performance using distance metrics, standard deviations, silhouette plots, and a comparison with groupings created by k-means clustering.

3 Methodology

3.1 Wellington and Auckland weather datasets

We use daily data for the 5-year period from 1 January 2017 to 31 December 2021, from Auckland and Wellington cities, collected from the Motat Ews weather station (Auckland) and the Kelburn Aws station (Wellington).

Data was downloaded from New Zealand's National Insitute for Water and Atmospheric Research (NIWA) National Climate Database (NIWA 2022b) on 9 August, 2022.

3.2 Variables

The dataset contains 7 variables of interest. These are:

- Station `(Nominal)`: the weather station at which the observation was made.
- Date(NZST) `(Date)`: the date on which the observation was made.
- Max Temperature `(Numeric)`: maximum temperature recorded in the 24 hour period. Units: degrees Celsius.
- Min Temperature `(Numeric)`: minimum temperature recorded in the 24 hour period. Units: degrees Celsius.
- Global Radiation `(Numeric)`: solar radiation incident on the earth's surface during the 24 hour period. This includes both radiation from direct sunlight and from diffuse (scattering) sources in the earth's atmosphere such as clouds. Units: megajoules per square metre.
- Wind `(Numeric)`: the total distance travelled by surface wind during the 24 hour period. Units: kilometres.
- Rainfall `(Numeric)`: the amount of precipitation within the 24 hour period. Units: millimetres.

We also derived the following variables from Date(NZST):

- Season `(Nominal)`: the season of the observation, based on the conventional four seasons:
  - Spring: September, October and November.
  - Summer: December, January and February.
  - Autumn: March, April and May.
  - Winter: June, July and August.
- Real Season `(Nominal)`: the alternative season of the observation, based on the 'Real NZ Weather' categorisation (Shand, A and Ivshin, D 2022). The seasons in this categorisation correspond to the following months
  - Spring 1_2: August and December
  - Shitsville: September, October and November (the same dates as the conventional spring season).
  - Summer: January and February.

  - Autumn: March, April and May (the same dates as the conventional autumn season).
  - Winter: June and July.

3.3 Exploratory data analysis

The methods used in our exploratory data analysis were:

- Tables of basic summary statistics (sample size, minimum, first quartile, median, third quartile, maximum, mean, standard deviation, skewness, and kurtosis).
- Boxplots of distribution of all numeric variables split by city and season (for both season and real season).
- Pairs plots showing densities of the distribution of varaibles, scatter plots, and correlation coefficients (for both season and real season).
- Tables of Anderson-Darling test for normality by season and city.
- Hotelling test for equality of means by season and city.
- Variance test for equality of variances season and city.

3.4 Comparison of the distances between seasonal summary statistics for the season and real season classifications

For both season and real season, and separately for Auckland and Wellington, we calculated seasonal summary statistics for each weather variable. This gave us a total of 29 values for each season: minimum(Tmax), first quartile(Tmax), median(Tmax), third quartile(Tmax), maximum(Tmax), mean(Tmax), minimum(Tmin), first quartile(Tmin), median(Tmin), third quartile(Tmin), maximum(Tmin), mean(Tmin), minimum(GlobalRad), first quartile(GlobalRad), median(GlobalRad), third quartile(GlobalRad), maximum(GlobalRad), mean(GlobalRad), minimum(WindRun), first quartile(WindRun), median(WindRun), third quartile(WindRun), maximum(WindRun), mean(WindRun), first quartile(log(Rainfall)), median(log(Rainfall)), third quartile(Rainfall), maximum(Rainfall), mean(Rainfall). Tables of these summary statistics are provided in Appendix A.

The log of rainfall was used instead of raw rainfall values where it was necessary because all of the summary statistics for rainfall equalled zero. Each weather variable has 6 summary statistics that were used in distance calculations, with the exception of rainfall/log(rainfall), where the minimum was omitted because it showed no variation across seasons.

Thus we obtained 4 new vectors for the season classification and 5 for the real season classification. Given:

Each weather variable in:
$W = \{Tmax,\ Tmin,\ GlobalRad,\ WindRun,\ Rainfall\ or\ log(Rainfall)\}$

And each summary statistic in:
$S = \{minimum,\ first\ quartile,\ median,\ third\ quartile,\ maximum,\ mean\}$

For the season classification we obtained the 4 vectors:

$$X_{Spring} = \{S_1(W_1), S_2(W_2), S_3(W_3), ...S_6(W_5)\}$$
$$X_{Summer} = \{S_1(W_1), S_2(W_2), S_3(W_3), ...S_6(W_5)\}$$
$$X_{Autumn} = \{S_1(W_1), S_2(W_2), S_3(W_3), ...S_6(W_5)\}$$
$$X_{Winter} = \{S_1(W_1), S_2(W_2), S_3(W_3), ...S_6(W_5)\}$$

(with the exception of the minimum of $Rainfall\ or\ log(Rainfall)$)

And for the real season classification we obtained the 5 vectors:

$$X_{Spring1\_2} = \{S_1(W_1), S_2(W_2), S_3(W_3), ...S_6(W_5)\}$$
$$X_{Shitsville} = \{S_1(W_1), S_2(W_2), S_3(W_3), ...S_6(W_5)\}$$
$$X_{Summer} = \{S_1(W_1), S_2(W_2), S_3(W_3), ...S_6(W_5)\}$$
$$X_{Autumn} = \{S_1(W_1), S_2(W_2), S_3(W_3), ...S_6(W_5)\}$$
$$X_{Winter} = \{S_1(W_1), S_2(W_2), S_3(W_3), ...S_6(W_5)\}$$

(with the exception of the minimum of $Rainfall\ or\ log(Rainfall)$)

For each seasonal classification and city, we calculated the Canberra distances between the seasons (i.e. between vectors X_{Spring}, X_{Summer}, X_{Autumn}, and X_{Winter} for season, and between $X_{Spring1\_2}$, $X_{Shitsville}$, X_{Summer}, X_{Autumn}, and X_{Winter} for real season).

The Canberra distance between seasons i and j is computed as:

$$d_C(X_i, X_j) = \sum_{\ell=1}^{29} \frac{| x_\ell - y_\ell |}{| x_\ell | + | y_\ell |}$$

We chose the Canberra distance because its denominator provides a form of scaling that prevents variables with larger values and variances (such as the WindRun summary statistics) from dominating the calculated distances. An alternative would be to normalise summary statistics for each weather variable by calculating their Z-scores. However, we expected (and found) that standardisation to zero mean and unit variance almost completely eliminated the overall differences in distances between the seasonal classifications, meaning that we could not use the average distance between seasons (computed using normalised data) to compare the classifications. The Canberra distance provides a good solution, levelling out the influence of different weather variables, but not forcing all summary statistics to have the same variance.

We compared the Canberra distances between the conventional and real season classifications. The rationale is that larger distances should be found for the classification that more effectively separates days into groups with dissimilar weather values.

3.5 Comparison of the standard deviations of the season and real season weather variables

For both season and real season, and separately for Auckland and Wellington, we calculated the standard deviations of each weather variable by season. The standard deviations for the season and the real season classifications were compared, with the rationale that lower standard deviations should be found for the classification that more effectively groups together days with similar weather values.

All standard deviations can be found in the summary statistic tables in Appendix A.

3.6 Comparison of silhouette plots for the season and real season classifications

We constructed silhouette plots from standardised daily weather data for season and real season and separately for Wellington and Auckland. The weather variables used in these plots were Tmax, Tmin, GlobalRad, WindRun, and Rainfall. We did not use log values for rainfall. Euclidean distances were used to construct the silhouette coefficients for each day and the plots display each day's silhouette coefficient.

The silhouette coefficient indicates how similar the day's observation is to the representative point for the season to which it is assigned, versus its similarity to the representative point of the next closest season. Positive values indicate that the day has been assigned to the season it is most similar to, while negative values indicate that the day is more similar to a different season. Where $a_{(i)}$ is the distance from the observation to all other observations in its season and $b_{(i)}$ is the distance from the observation to all observations in the next closest season, the silhouette coefficient ($s_{(i)}$) is calculated as:

$$s_{(i)} = \frac{b_{(i)} - a_{(i)}}{max(b_{(i)}, a_{(i)})}$$

3.7 Comparison of k-means clusters to the season and real season classifications

We used k-means clustering to create new weather-based classifications of daily data into 4 and 5 groups. The clustering used standardised daily weather data and was done separately for Wellington and Auckland. We used the Hartigan-Wong k-means clustering algorithm with 50 randomly assigned start points and 100 maximum iterations.

We investigated the quality of the k-means clusters using cluster plots and silhouette plots. We then assessed the performance of the conventional season classification against the 4 k-means clusters and the real season classification against the 5 k-means clusters. This follows the rationale that if the k-means clusters represent an optimal clustering of days with similar weather, then the season classification that is most similar to the k-means clusters may be a better classifier of the weather.

Once the k-means clusters were produced, the next task was to determine which season should be matched to which cluster. We chose the best match as the one that gave the highest number of correct predictions.

For the best match, we then computed confusion matrices and assessed the performance of each seasonal classification relative to the k-means clusters using overall accuracy, macro precision, macro recall, and macro F1 scores. In the description of these measures below, a true positive value (TP_i) is where the seasonal classification correctly predicts that an observation is a member of the class, a true negative (TN_i) is where the seasonal classification correctly predicts that an observation is not a part of the class, a false positive (FP_i) is where the seasonal classification incorrectly predicts that an observation is part of the class, and a false negative (FN_i) is where the seasonal classification incorrectly predicts that an observation is not part of the class.

- **Overall accuracy:** the overall proportion of correct predictions, across all classes.
- **Macro precision:** a measure of the proportion of predictions, for each class, that are correct. This measure averages the precision values across classes, where precision is calculated as:

  $PPV_i = \frac{TP_i}{TP_i + FP_i}$
- **Macro recall:** a measure of the proportion of observations that are of the k-means class, that were correctly predicted as such by the seasonal classification. This is the average of the recall values across classes, where recall is calculated as:

  $TPR_i = \frac{TP_i}{TP_i + FN_i}$
- **Macro F1:** a measure of the harmonic mean of precision and recall, which is a better measure of overall accuracy in situations where classes are unbalanced. This is computed as the average of the F1 scores across the classes, where the F1 score is calculated as:

  $F1\ score_i = \frac{2(PPV_i \times TPR_i)}{PPV_i + TPR_i}$

3.8 Software

All computations were performed using the statistical software R. We used the R packages: dplyr, tidyr, lubridate, moments, ggplot2, GGally, nortest, Hotelling, grid, gridExtra, pheatmap, viridis, RColorBrewer, cluster, factoextra, vegan, and caret.

4 Results

4.1 Exploratory data analysis

4.1.1 Summary statistics by city and season

Tables of summary statistics for daily weather data by season and city are provided in Appendix A

Our main observations based on these summary statistics are as follows.

- The rainfall for both Wellington and Auckland appears to have a heavy right skew with the minimum through to median value often being 0 for all seasons regardless of city, and then the max value being significantly larger.
- Wellington has much higher median WindRun for all seasons compared to Auckland.
- Auckland has a higher median Tmax value for all seasons compared to the equivalent season for Wellington.
- Auckland also has a higher median GlobalRad value for all seasons compared to the equivalent season for Wellington.
- The skewness and kurtosis values for nearly all variables for all seasons indicate that most of these variables may not be normally distributed.
- The median values for all weather variables between Spring1_2 and Shitsville for the realseason classification are all very similar to one another, which isn't a great sign for this system being more accurate.

4.1.2 Boxplots of weather variables by city by season

Figure 3 shows boxplots of the distribution of each variable by city and conventional season. From these plots we draw the following conclusions.

Global radiation

- For global radiation, summer has the highest average and median values and winter the lowest. Global radiation is a measure of direct and diffuse sunlight, so this is consistent with changes in day length across seasons.
- The distribution of global radiation values is more spread out in summer and has the lowest spread in winter. This may relate to cloud cover, with the longer summer days giving rise to more variation because there is more time for cloud cover to have an effect.
- There are some high outlier days for global radiation in both cities in winter, and some low outliers for Auckland in summer.

Rainfall

- log(Rainfall) appears very similarly distributed across cities and seasons, although there may be more variablity in Wellington's rainfall in autumn and winter and more variability for Auckland in summer.

Maximum temperature

- Across every season, Auckland has higher maximum temperatures. The bottom of the interquartile range of Auckland's maximum temperatures is level with the top of the interquartile range of Wellington's maximum temperatures, suggesting that 75% of Auckland's days reach maximum temperatures that are only reached by 25% of Wellington's days.
- There are some high and low outlier values for maximum temperature in both cities and during all seasons.

Minimum temperature

- Minimum temperatures are less different between Wellington and Auckland than maximum temperatures, but mean and median minimum temperatures are somewhat lower in Wellington than in Auckland.

Wind run

- Wind run shows the most dramatic difference between Wellington and Auckland. Mean and median values in Wellington are approximately twice those of Auckland, and around 75% of Wellington's days recorded wind run values that were reached by only 25% of Auckland's days.
- Wellington also shows more variability in wind run values than Auckland.
- Wellington and Auckland both have some high outlier values for wind run, reflecting some very windy days. However Wellington's high outliers are much higher than Auckland's.

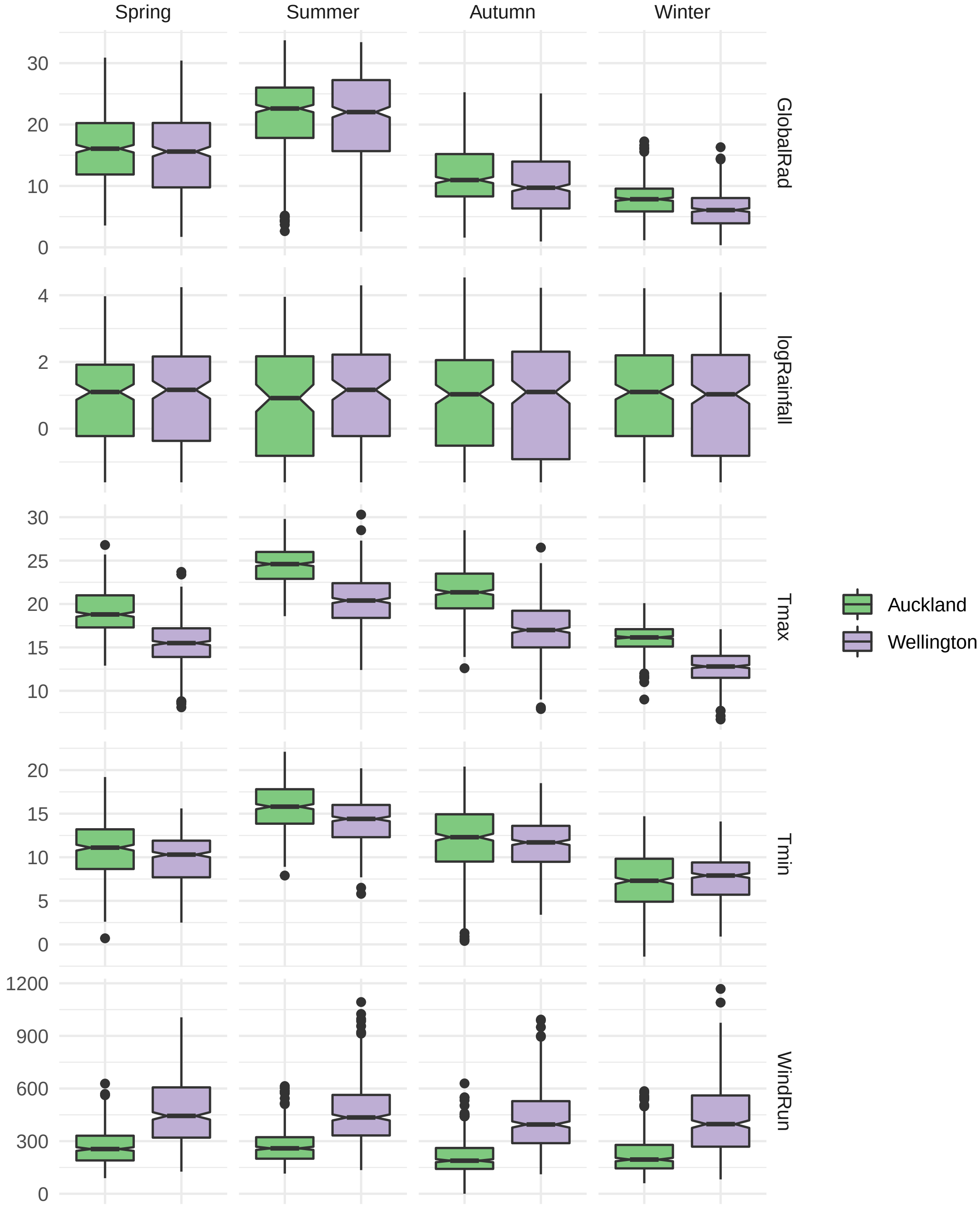


Figure 3: Boxplots of the distribution of each weather variable by city and season. The units of each weather variable are given in the Methods section. Data is pooled for the 5 year period: 2017-2021.

4.1.3 Boxplots of weather variables by city and real season

Figure 4 shows boxplots of the distribution of each variable by city and season. From these plots we see the following key differences compared to conventional seasons.

- In both Auckland and Wellington, Spring1_2 and Shitsville appear to have very similar weather patterns.
- We do not see evidence for higher rainfall or wind during Shitsville as compared to Spring1_2.

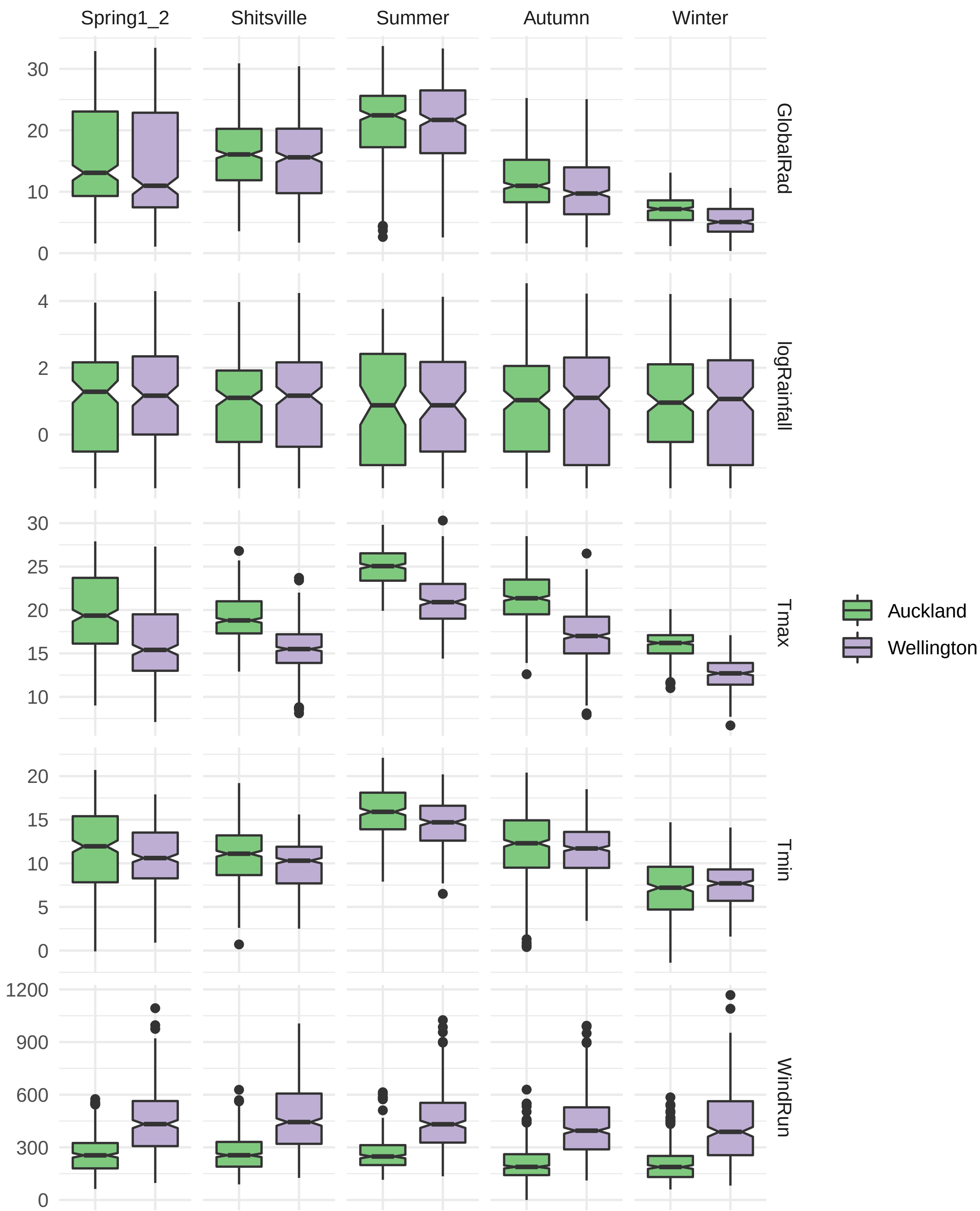


Figure 4: Boxplots of the distribution of each weather variable by city and the alternative 'real season' categorisation. The units of each weather variable are given in the Methods section. Data is pooled for the 5 year period 2017-2021.

4.1.4 Pairs plot of the relationships between weather variables by season

We draw the following main conclusions from the pairs plot for conventional seasons (Figure 5).

Maximum and minimum temperatures

- Maximum and minimum temperatures are strongly (and significantly) positively correlated during all seasons, although the correlation may be more moderate in winter.
- The relationship between maximum and minimum temperature appears (from the scatter plot) to be linear.

Relationships between temperature and global radiation

- Maximum temperature is positively correlated with global radiation during all seasons except summer, consistent with sunshine being a driver of higher temperatures. We are uncertain why this relationship is not apparent in summer.
- Minimum temperature has a weak (but significant) negative correlation with global radiation in summer and winter, but an overall positive correlation when aggregated across seasons. This was unexpected, but it may indicate that:
  - when data is pooled across all seasons, higher temperatures and more sunshine are generally found together due to seasonal effects.
  - when season is allowed for, sunny clear days (with higher global radiation values) are associated with clear nights, which tend to be colder due to the absence of the insulating effect of clouds.

Relationships between temperature and wind

- There is a moderate to strong negative correlation between maximum temperature and wind run across all seasons.
- The relationship between minimum temperature and wind run is not consistent across seasons, with the only clear result being a significant positive correlation in winter.

Relationships between wind and global radiation

- Similar to wind and maximum temperature, we see negative correlations between global radiation and wind. The correlations are moderate to weak, but they are consistent across all seasons.

Relationships between rainfall and other variables

- Rainfall shows weak but significant negative correlations with global radiation, across all seasons. This is consistent with the idea that rain requires cloud cover, which also reduces radiation from the sun.
- Rainfall is negatively correlated with maximum temperatures but none of these correlations are significant.

- The only convincingly significant correlation between rainfall and minimum temperature is in winter, where it is positively correlated with higher minimum temperatures. This is consistent with the interpretation that cloudy nights are somewhat warmer.
- Rainfall and wind run are positively correlated in all seasons.
- In the scatterplots, log(Rainfall) appears non-continuous at lower values. This is a combination of the effect of the log transformation on the lowest values, and the fact that rainfall measurements are rounded to the nearest 0.2mm (see the section above on Further investigation of the rainfall variable).

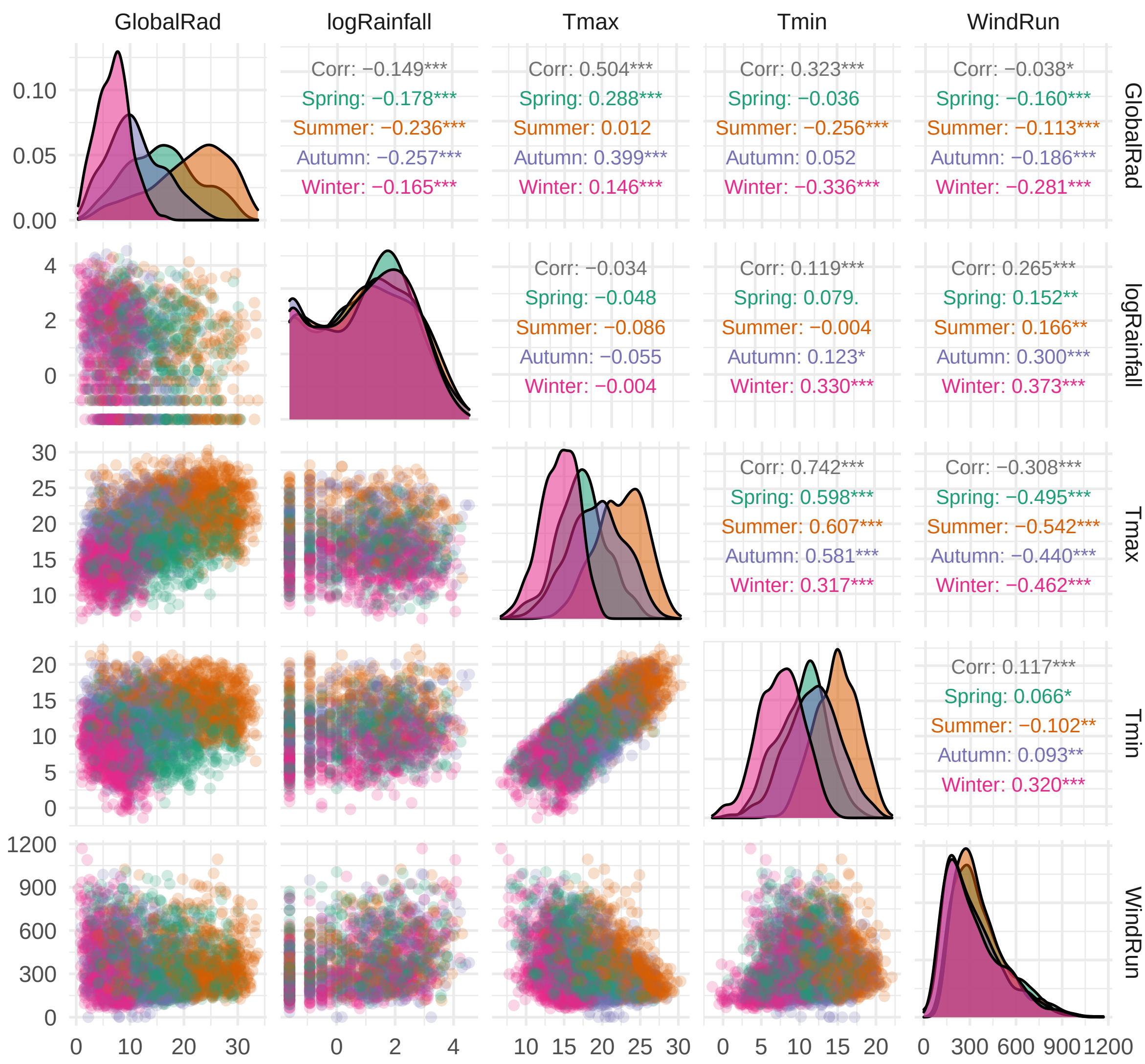


Figure 5: Pairs plot showing the relationships between the weather variables for spring (teal), summer (orange), autumn (purple), and winter (pink). Two dimensional scatterplots for each pair of variables are in the lower left panels, with observations colour-coded by season. A histogram of the variable distributions by season is shown in the diagonal, and correlation coefficients are shown in the upper right panels. Asterisks indicate the significance levels of the correlaion coefficients, with *** indicating $p < 0.001$, ** $p < 0.01$, and * $p < 0.05$. Units of measurement are described in the Methods section of this report.

4.1.5 Pairs plot of the relationships between weather variables by real season

As shown in Figure 6, we see the following main differences in the correlations between weather variables when we compare Spring1_2 and Shitsville.

- Global radiation and maximum temperature appear to be more strongly positively correlated in Spring1_2 than in Shitsville. This might relate to the presence of longer December days in the Spring1_2 period.
- We see a significant moderate positive correlation between minimum temperatures and global radiation during spring1_2 but not during shitsville.
- The positive correlation between minimum and maximum temperature is stronger in spring1_2 than in shitsville.
- There is a moderate negative correlation between wind run and maximum temperature during shitsville. While this correlation is also present (and significant) during spring1_2, it is weaker.

It is difficult to speculate on the causes of these differences, but at least some may relate to the more variable day lengths in Spring1_2 than Shitsville.

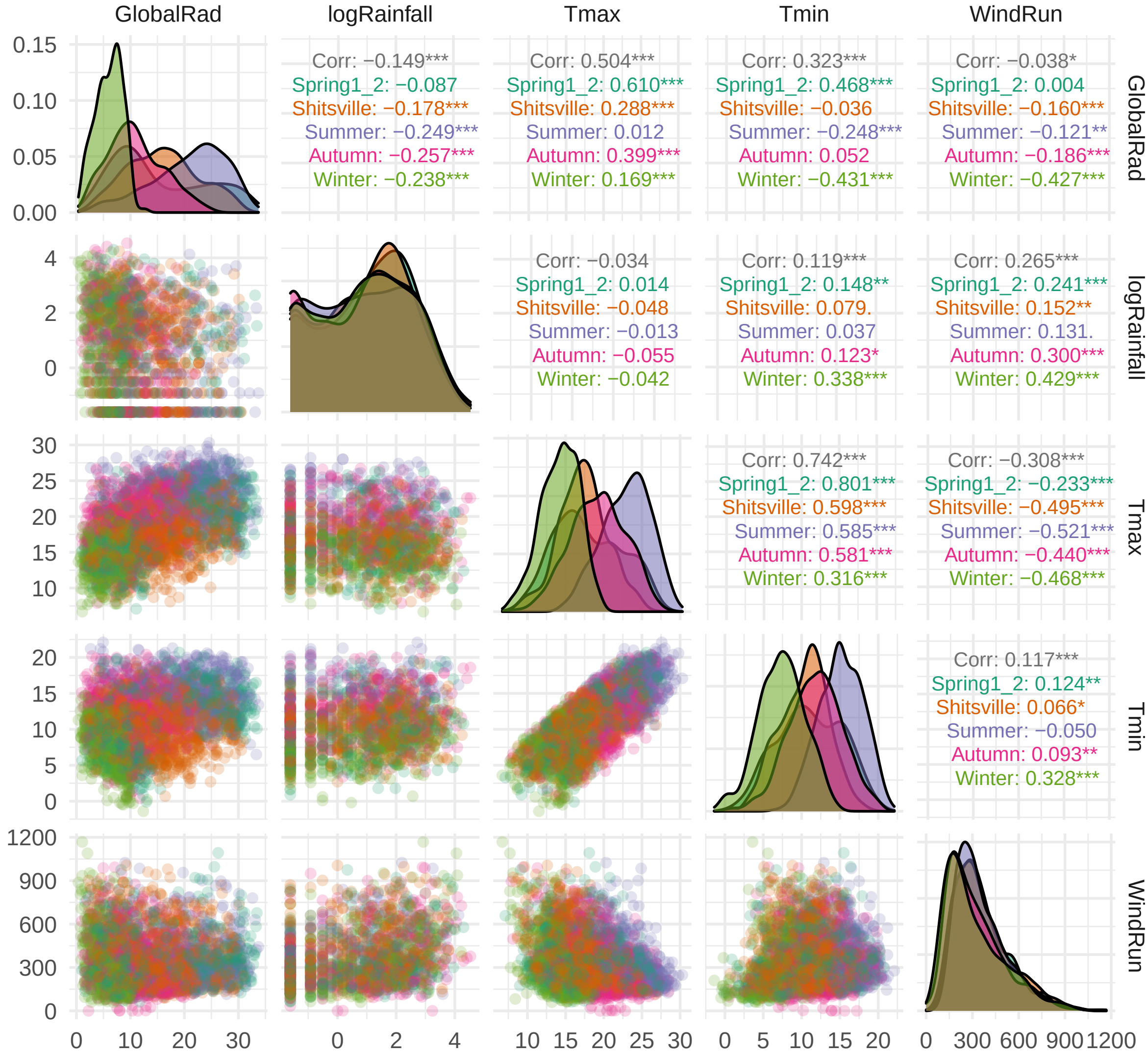


Figure 6: Pairs plot showing the relationships between the weather variables for spring 1 and 2 (teal), shitsville (orange), summer (purple), and autumn (pink), and winter (green). Two dimensional scatterplots for each pair of variables are in the lower left panels, with observations colour-coded by season. A histogram of the variable distributions by season is shown in the diagonal, and the correlation coefficients are shown in the upper right panels. Asterisks indicate the significance levels of the correlaion coefficients, with *** indicating p < 0.001, ** indicating p < 0.01, and * indicating p < 0.05. Units of measurement are described in the Methods section of this report.

4.1.6 Anderson-Darling test of Normality

We used the Anderson Darling test of normality to determine whether each weather variable, in each city and each season, is normally distributed. This is in preparation for further analysis where non-normality can influence the choice of method, the conclusions that can be drawn from different methods and decisions to transform non-normal variables. The null hypothesis for this test is that the data being tested is normally distributed. If the p-value for this test is > 0.05 then we do not have evidence to reject the null hypothesis of normality at the 5% significance level.

Table 1 presents the results of Anderson-Darling tests for the conventional season classification in Wellington.

We *don't have enough evidence to reject* normality at the 5% significance level for the following variables for Wellington by season:

- Spring: Tmax
- Summer: Tmax
- Autumn: Tmax, Tmin

All other variables appear to be non-normally distributed.

Table 1: Table of p-values from Anderson Darling test of normality for Wellington by season for all variables

	Tmax	Tmin	GlobalRad	WindRun	Rainfall
Spring	0.064034	0.000000	0.000190	0	0
Summer	0.517325	0.008857	0.000000	0	0
Autumn	0.535481	0.057139	0.000000	0	0
Winter	0.000247	0.004812	0.001943	0	0

Table 2 presents the results of Anderson-Darling tests for the conventional season classification in Auckland.

We find that we *don't have enough evidence to reject* normality at the 5% significance level for the following variables for Auckland by season:

- Autumn: Tmax, Tmin

All other variables appear to be non-normally distributed.

Table 2: Table of p-values from Anderson Darling test of normality for Auckland by season for all variables

	Tmax	Tmin	GlobalRad	WindRun	Rainfall
Spring	0.000020	0.006933	0.000148	0	0
Summer	0.027852	0.008778	0.000000	0	0
Autumn	0.066280	0.237327	0.000017	0	0
Winter	0.000247	0.004812	0.001943	0	0

Table 3 presents the results of Anderson-Darling tests for the real season classification in Wellington.

We *don't have enough evidence to reject* normality at the 5% significance level for the following variables for Wellington by real season:

- Shitsville: Tmax
- Summer: Tmax
- Autumn: Tmax, Tmin
- Winter: Tmin

All other variables appear to be non-normally distributed.

Table 3: Table of p-values from Anderson Darling test of normality for Wellington by real season for all variables

	Tmax	Tmin	GlobalRad	WindRun	Rainfall
Spring1_2	0.000000	0.007608	0.000000	2e-05	0
Shitsville	0.064034	0.000000	0.000190	0e+00	0
Summer	0.645945	0.018312	0.000000	2e-06	0
Autumn	0.535481	0.057139	0.000000	0e+00	0
Winter	0.016833	0.093431	0.000469	1e-06	0

Table 4 presents the results of Anderson-Darling tests for the real season classification in Auckland.

We *don't have enough evidence to reject* normality at the 5% significance level for the following variables for Auckland by real season:

- Summer: Tmax
- Autumn: Tmax, Tmin
- Winter: Tmax, Tmin

All other variables appear to be non-normally distributed.

Table 4: Table of p-values from Anderson Darling test of normality for Auckland by real season for all variables

	Tmax	Tmin	GlobalRad	WindRun	Rainfall
Spring1_2	0.000000	0.000010	0.000000	2.1e-05	0
Shitsville	0.000020	0.006933	0.000148	0.0e+00	0
Summer	0.166399	0.007745	0.000000	0.0e+00	0
Autumn	0.066280	0.237327	0.000017	0.0e+00	0
Winter	0.107218	0.199534	0.000021	0.0e+00	0

4.1.7 Tests for equality of means for season

We use Hotelling tests for the equality of means by season. The null hypothesis of this test is that the means for both the Wellington variable and the Auckland variable are equal. This means that if the p-value is greater than 0.05 we cannot reject the null hypothesis at the 5% significance level.

From Table 5 we conclude that *we do not have enough evidence* to reject the null hypothesis of equal means between equivalent variables for Auckland and Wellington by season at the 5% significance level for the following variables:

- Summer: GlobalRad, Rainfall
- Autumn: Rainfall
- Winter: Tmin, Rainfall

This means that we do not have evidence for differences between Wellington and Auckland for Summer mean GlobalRad and Rainfall, Autumn mean Rainfall, or Winter mean Tmin and Rainfall. For all other variables we do reject the null hypothesis, suggesting that there are significant differences in the seasonal weather patterns between Wellington and Auckland.

Table 5: Table of p-values from test to compare means between Wellington and Auckland by season for all variables

	Tmax	Tmin	GlobalRad	WindRun	Rainfall
Spring	0	0.000000	0.036889	0	0.043066
Summer	0	0.000000	0.159505	0	0.138003
Autumn	0	0.002772	0.000352	0	0.317296
Winter	0	0.064697	0.000000	0	0.412597

4.1.8 Tests for equality of variance for season

We use F-tests of equality of variances for the season classification, with results shown in Table 6. The null hypothesis for this test is that there is equal variance between the Wellington variable and the Auckland equivalent of that variable. This means that if the p-value is greater than 0.05 we cannot reject the null hypothesis at the 5% significance level.

Results in Table 6 suggest *we do not have enough evidence* to reject the null hypothesis of equal variance between equivalent variables for Auckland and Wellington by season at the 5% significance level for the following variables:

- Spring: Tmax
- Summer: Tmin
- Autumn: Rainfall
- Winter: GlobalRad

For the remaining variables we do reject the null hypothesis and conclude that variances are not equal. This may be a problem for some hypothesis testing.

It is interesting that none of the variables that have equal variance between Wellington and Auckland are the same for each season.

Table 6: Table of p-values from F-test to compare variances between Wellington and Auckland by season for all variables

	Tmax	Tmin	GlobalRad	WindRun	Rainfall
Spring	0.215084	0.000137	0.000585	0	0.000000
Summer	0.000000	0.392424	0.001064	0	0.000046
Autumn	0.044729	0.000000	0.030972	0	0.708362
Winter	0.002288	0.000000	0.560819	0	0.000004

4.1.9 Tests for equality of means for real season

Hotelling tests of equality of means for the real season classification (Table 7) show that *we do not have enough evidence* to reject the null hypothesis of equal means between equivalent variables for Auckland and Wellington by real season at the 5% significance level for the following variables:

- Spring1_2: Rainfall
- Summer: GlobalRad, Rainfall
- Autumn: Rainfall
- Winter: Rainfall

This means that we do not have evidence for differences between Wellington and Auckland for Spring1_2 mean Rainfall, Summer mean GlobalRad and Rainfall, Autumn mean Rainfall, and Winter mean Rainfall. For all other variables we do reject the null hypothesis, suggesting that there are significant differences in the seasonal weather patterns between Wellington and Auckland when using the real season categorisation.

Table 7: Table of p-values from test to compare means between Wellington and Auckland by real season for all variables

	Tmax	Tmin	GlobalRad	WindRun	Rainfall
Spring1_2	0	0.006857	0.048898	0	0.261300
Shitsville	0	0.000000	0.036889	0	0.043066
Summer	0	0.000000	0.434605	0	0.834446
Autumn	0	0.002772	0.000352	0	0.317296
Winter	0	0.043905	0.000000	0	0.203551

4.1.10 Tests for equality of variance for real season

F-tests for equality of variances in the real season classification (Table 8) show that *we do not have enough evidence* to reject the null hypothesis of equal variance between equivalent variables for Auckland and Wellington by season at the 5% significance level for the following variables:

- Spring1_2: Tmax, GlobalRad
- Shitsville: Tmax
- Summer: Tmin, Rainfall
- Autumn: Rainfall
- Winter: GlobalRad

For the remaining variables we do reject the null hypothesis and conclude that variances are not equal. This may be a problem for some hypothesis testing.

It is interesting that the variables for which we don't reject the null hypotheses of equal variance between Wellington and Auckland for are very similar for realseason as the variables identified for season, with Summer only gaining the Rainfall variable, and Spring1_2 and Shitsville from realseason having very similar variables to Spring for season.

Table 8: Table of p-values from F-test to compare variances between Wellington and Auckland by real season for all variables

	Tmax	Tmin	GlobalRad	WindRun	Rainfall
Spring1_2	0.606611	0.000013	0.082113	0	0.000000
Shitsville	0.215084	0.000137	0.000585	0	0.000000
Summer	0.000001	0.311038	0.022888	0	0.603533
Autumn	0.044729	0.000000	0.030972	0	0.708362
Winter	0.003284	0.000000	0.772222	0	0.000059

4.2 Comparison of the distances between seasonal summary statistics for season and real season

As described in the Methodology, for both season and real season, and separately for Auckland and Wellington, we calculated seasonal summary statistics for each weather variable. We then used the summary statistics as a new dataset, in which each observation was a season and each variable was a summary statistic. We calculated Canberra distances between the observations (seasons). Below we compare the distances obtained from the season and real season classifications. We postulate that the classification that generates greater overall distances between seasons is a more effective separator of days into groups with dissimilar weather values.

4.2.1 Overall distances between seasons are greater for the season classification than the real season classification, for both cities

We find that, for both Wellington and Auckland, the average distance between seasons is greater for the conventional season classification (Tables 9 and 11) than for the real season classification (Tables 10 and 12). The difference is more pronounced for Auckland than for Wellington. For Wellington, the average Canberra distance between seasons is 5.25 for conventional seasons and only slightly lower, at 5.19 for real seasons. For Auckland, the average Canberra distance between seasons is 6.32 for conventional seasons and 5.90 for real seasons. While these differences between classifications are only moderate, they suggest that conventional seasons may be a slightly better fit with observed weather patterns, for both cities.

Table 9: Canberra distances for the conventional season classification, Wellington

	Canberra distance
Spring vs Summer	4.053
Spring vs Autumn	2.868
Spring vs Winter	5.730
Summer vs Autumn	4.973
Summer vs Winter	9.003
Autumn vs Winter	4.889
Average distance	5.253

Table 10: Canberra distances for the real season classification, Wellington

	Canberra distance
Spring1_2 vs Shitsville	3.002
Spring1_2 vs Summer	6.092
Spring1_2 vs Autumn	3.124
Spring1_2 vs Winter	6.315
Shitsville vs Summer	4.612
Shitsville vs Autumn	2.868
Shitsville vs Winter	6.045
Summer vs Autumn	5.192
Summer vs Winter	9.438
Autumn vs Winter	5.175
Average distance	5.186

Table 11: Canberra distances for the conventional season classification, Auckland

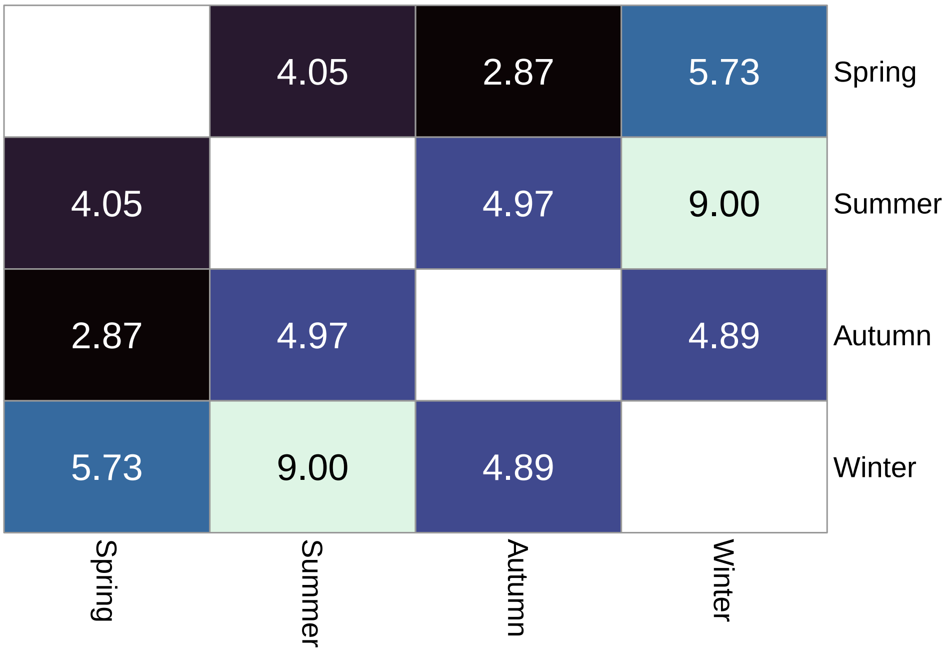
	Canberra distance
Spring vs Summer	4.856
Spring vs Autumn	4.341
Spring vs Winter	6.202
Summer vs Autumn	6.613

	Canberra distance
Summer vs Winter	9.462
Autumn vs Winter	6.466
Average distance	6.323

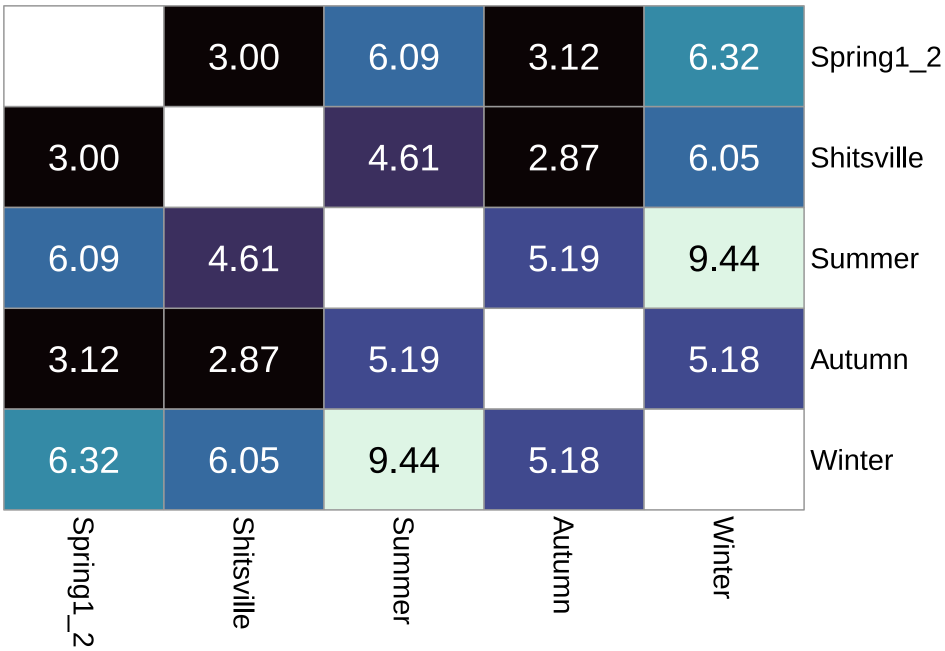
Table 12: Canberra distances for the real season classification, Auckland

	Canberra distance
Spring1_2 vs Shitsville	3.370
Spring1_2 vs Summer	5.697
Spring1_2 vs Autumn	4.397
Spring1_2 vs Winter	6.002
Shitsville vs Summer	5.235
Shitsville vs Autumn	4.341
Shitsville vs Winter	6.514
Summer vs Autumn	6.944
Summer vs Winter	9.968
Autumn vs Winter	6.579
Average distance	5.905

A comparison of the distances between specific seasons shows that, for conventional seasons, the largest distances are between summer and winter (as we would expect) and the smallest distances are between spring and autumn and between spring and summer (Figures 7a and 8a). The real season classification also has its largest distances between summer and winter. Notably, given that a central tenet of the real season classification is that there is a difference between Spring1_2 and Shitsville weather, we see only small distances between these two seasons (Figures 7b and 8b). This suggests that Spring1_2 and Shitsville may not have especially different weather patterns.

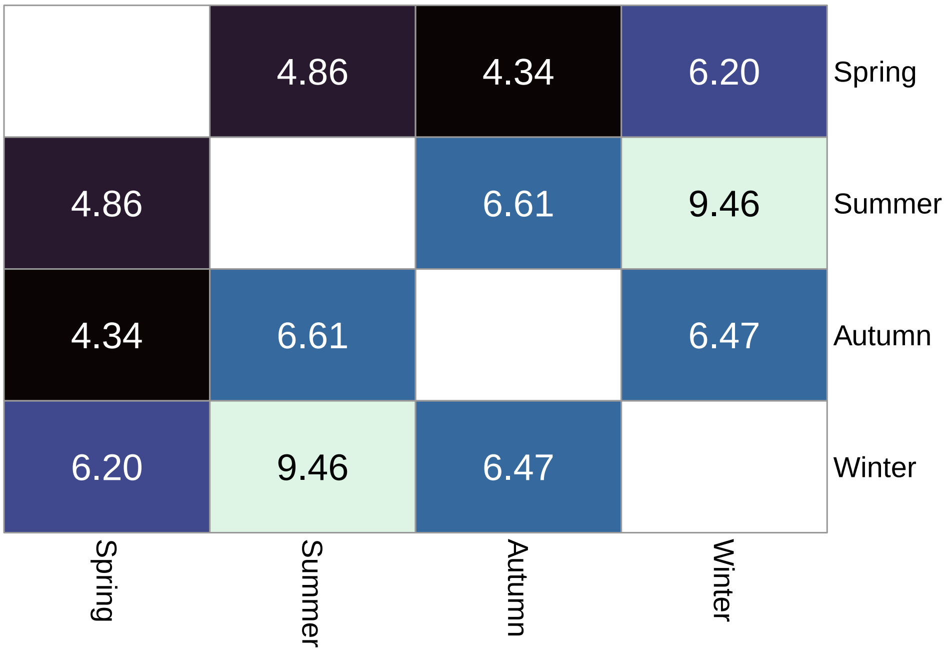


(a) conventional seasons

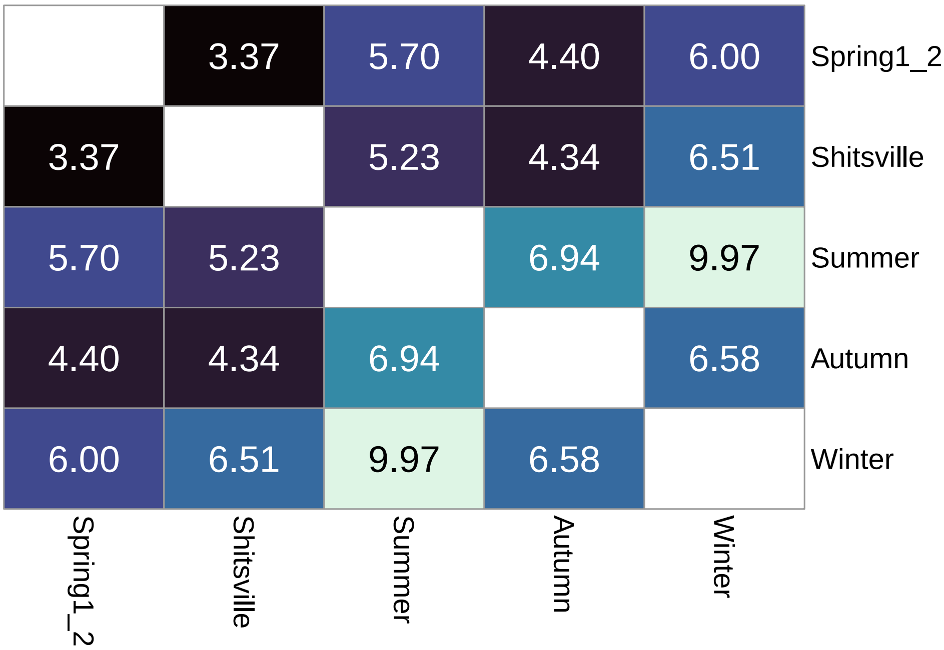


(b) real seasons

Figure 7: Heatmaps of the Canberra distances computed between each season for Wellington, comparing the season classifications. Darker colours correspond to shorter distances and lighter colours to longer distances.



(a) conventional seasons



(b) real seasons

Figure 8: Heatmaps of the Canberra distances computed between each season for Auckland, comparing the season classifications. Darker colours correspond to shorter distances and lighter colours to longer distances.

4.3 Comparison of the standard deviations of the season and real season weather variables

We computed the standard deviations of each weather variable, for each season and city (see Methodology). Below we compare the standard deviations obtained for the season and real season classifications. The classification that groups together days with more similar weather values should have lower standard deviations.

4.3.1 For both cities, most standard deviations are smaller for the season classification than for real season

When the standard deviations for each weather variable are averaged across seasons, we find that most averages are smaller for the conventional season classification than for the real season classification, both for Wellington (Tables 13 and 14) and Auckland (Tables 15 and 16). The exceptions to this are the average standard deviations for rainfall in Wellington and Auckland, and for global radiation and wind run in Auckland, where we see higher average standard deviations for the conventional season classification. However, average standard deviation differences between the classifications are not especially pronounced.

Of note is that, for both cities, the standard deviations of all weather variables except wind run are higher for the real season Spring1_2 than for the conventional season of Spring. The Spring1_2 standard deviations are also among some of the highest standard deviations seen for any of the seasons in the real season classification. This suggests that Spring1_2 may be grouping together days with dissimilar weather. This is not especially surprising for the global radiation value, given that Spring1_2 groups together two parts of the year where the day lengths are different, but it is interesting that it also holds true for other weather variables.

Overall, these results suggest that the conventional season classification may be a slightly better fit with observed weather patterns.

Table 13: Standard deviations for the conventional season classification, Wellington

season	sd(Tmax)	sd(Tmin)	sd(GlobalRad)	sd(WindRun)	sd(Rainfall)
Summer	2.85	2.62	7.55	174.05	8.10
Autumn	3.07	2.96	5.37	178.15	8.76
Winter	1.88	2.38	2.99	198.46	9.07
Spring	2.75	2.82	7.11	185.03	7.92
Average across seasons	2.64	2.69	5.75	183.92	8.46

Table 14: Standard deviations for the real season classification, Wellington

season	sd(Tmax)	sd(Tmin)	sd(GlobalRad)	sd(WindRun)	sd(Rainfall)
Summer	2.82	2.67	7.21	171.22	6.33
Autumn	3.07	2.96	5.37	178.15	8.76

season	sd(Tmax)	sd(Tmin)	sd(GlobalRad)	sd(WindRun)	sd(Rainfall)
Winter	1.94	2.39	2.36	208.62	9.70
Spring1_2	3.99	3.66	9.08	179.37	9.24
Shitsville	2.75	2.82	7.11	185.03	7.92
Average across seasons	2.91	2.90	6.22	184.48	8.39

Table 15: Standard deviations for the conventional season classification, Auckland

season	sd(Tmax)	sd(Tmin)	sd(GlobalRad)	sd(WindRun)	sd(Rainfall)
Summer	2.19	2.73	6.47	88.88	6.68
Autumn	2.80	3.91	4.85	97.35	8.92
Winter	1.63	3.35	2.91	104.55	7.30
Spring	2.59	3.37	6.03	97.39	6.10
Average across seasons	2.30	3.34	5.07	97.04	7.25

Table 16: Standard deviations for the real season classification, Auckland

season	sd(Tmax)	sd(Tmin)	sd(GlobalRad)	sd(WindRun)	sd(Rainfall)
Summer	2.12	2.83	6.31	90.14	6.52
Autumn	2.80	3.91	4.85	97.35	8.92
Winter	1.64	3.51	2.32	102.84	7.69
Spring1_2	4.11	4.69	8.22	97.18	6.81
Shitsville	2.59	3.37	6.03	97.39	6.10
Average across seasons	2.65	3.66	5.55	96.98	7.21

4.4 Comparison of silhouette plots for the season and real season classifications

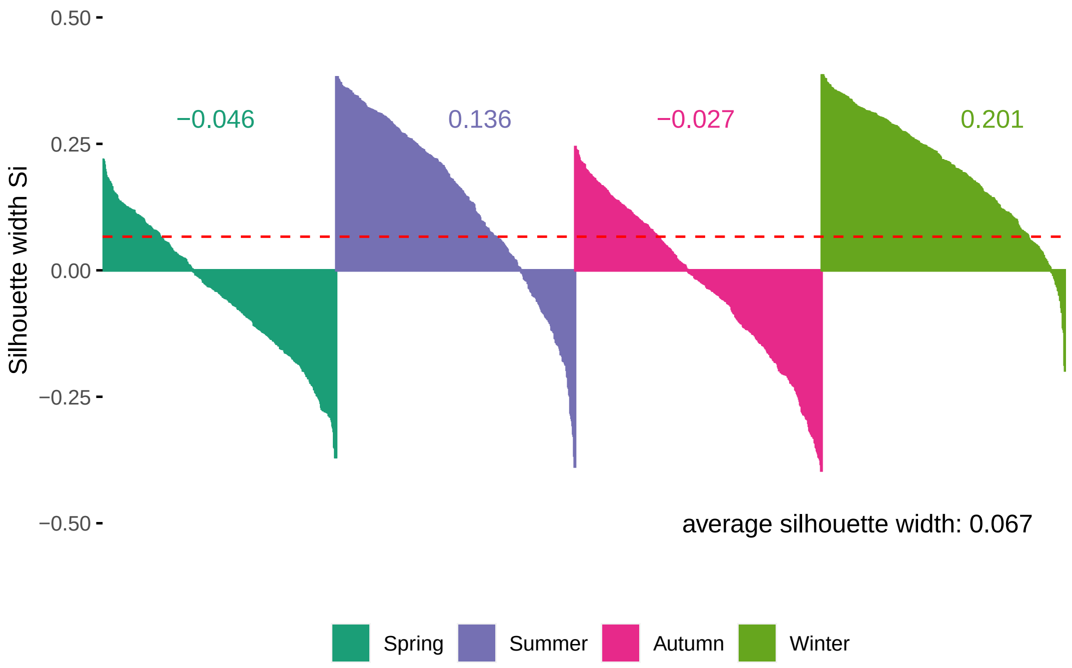
We compared the season and real season classifications using silhouette plots of standardised daily weather data, as described in the Methodology. We reason that the classification that more effectively groups together days with similar weather will have higher silhouette coefficient values, on average.

4.4.1 For both cities, average silhouette widths are higher for the season classification than for real season

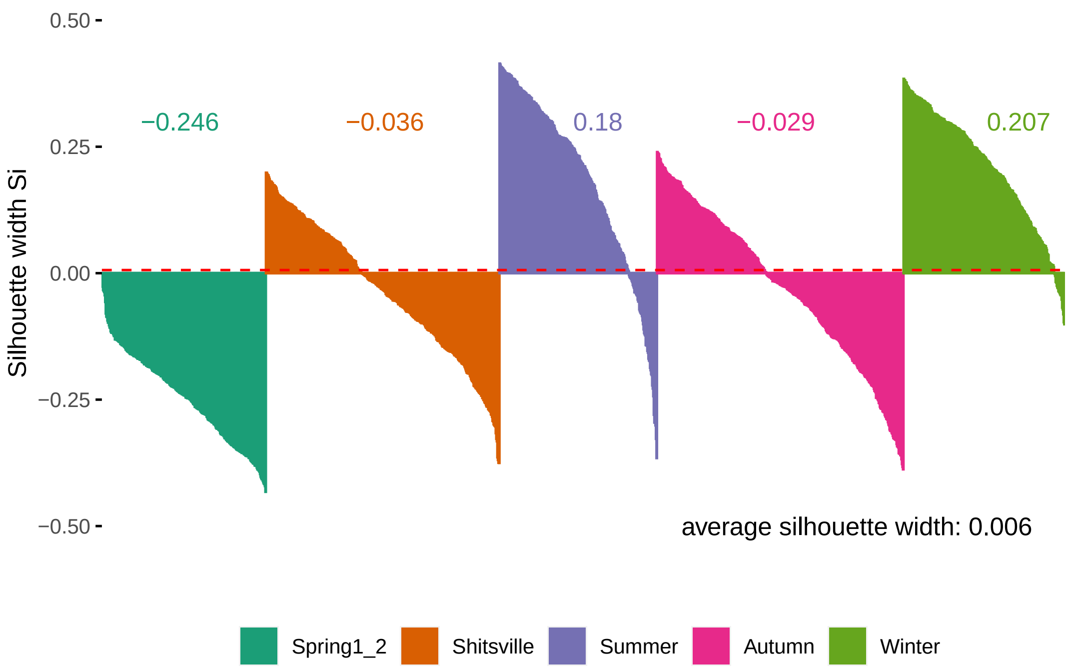
Figure 9 compares the silhouette plots between season and real season for Wellington. Figure 10 makes the same comparison for Auckland.

For both cities we see that the average silhouette width is higher for the conventional season classification than for real season, suggesting that the conventional seasons are better at grouping together days with similar weather. Notably for the real season classification, is that every silhouette coefficient in Spring1_2 is negative, indicating that every single day would be more appropriately classified into a different season (Figures 9b and 10b).

Interestingly, the silhouette plots also indicate that neither seasonal classification does an especially good job of classifying together days with similar weather. The classifications perform reasonably well for summer and winter, with positive silhouette coefficients for most days and silhouette widths ranging from 0.13 to 0.25. Spring and autumn perform poorly, however, with negative silhouette coefficients for most days and silhouette widths below zero for both cities (Figures 9a and 10a).

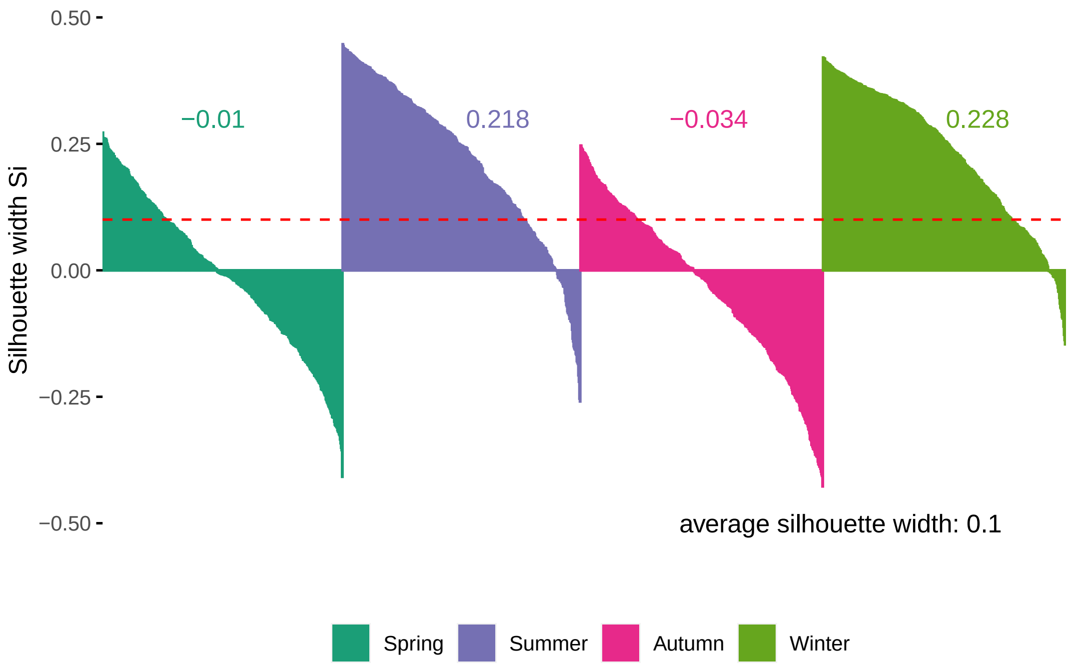


(a) conventional seasons

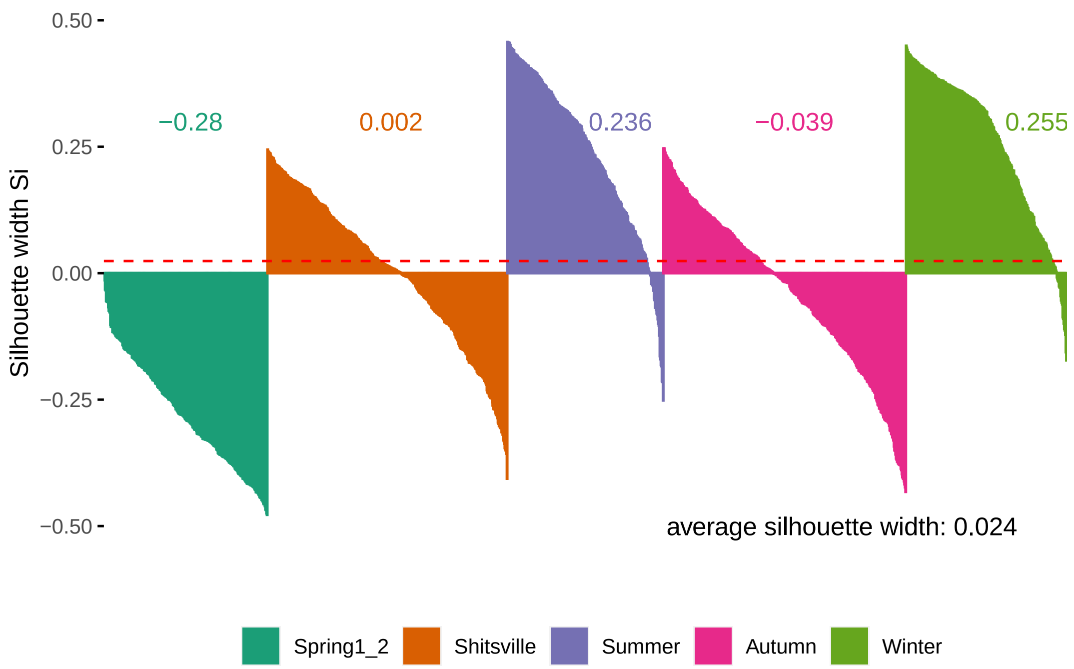


(b) real seasons

Figure 9: Silhouette plots for Welllington showing the silhouette widths for standardised daily data from each season, comparing the two season classifications. The silhouette width for each season is displayed in the text above each season's values and the average silhouette width across all seasons is indicated by the red dashed line and the text at the bottom of the chart



(a) conventional seasons



(b) real seasons

Figure 10: Silhouette plots for Auckland showing the silhouette widths for standardised daily data from each season, comparing the two season classifications. The silhouette width for each season is displayed in the text above each season's values and the average silhouette width across all seasons is indicated by the red dashed line and the text at the bottom of the chart

4.5 Comparison of kmeans clustering with the season and real season classifications

We performed k-means clustering of standardised daily weather data from each city (specifying 4 and 5 groups) to create new weather-based classifications. We then assessed the performance of the conventional season classification against the 4 k-means clusters and the performance of the real season classification against the 5 k-means clusters, as described in the Methodology.

Our rationale is that the k-means clusters may represent a more an optimal clustering of days with similar weather than the seasonal classifications. If this is the case, then the seasonal classification that is most similar to the k-means clusters may be a better classifier of the weather.

4.5.1 K-means clustering of Wellington daily weather data

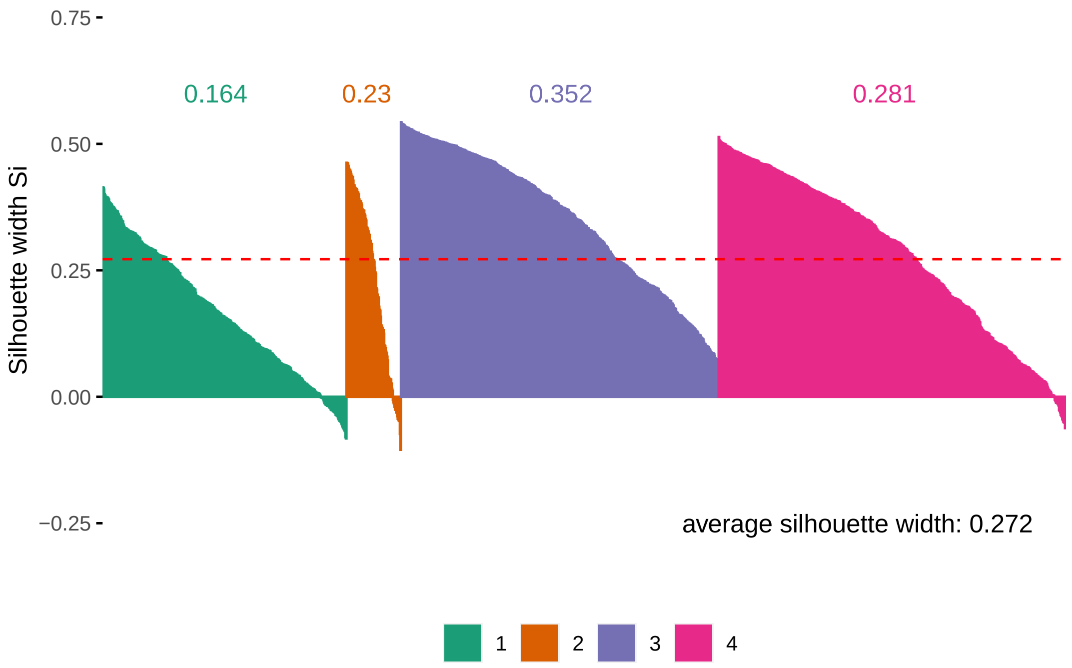
K-means clustering of standardised daily Wellington weather data does appear to generate clusters with greater in-group similarity than either of the two seasonal classifications. Silhouette plots show that the silhouette widths of the four and five k-means clusters (Figure 12) are consistently larger than those of the season and real season classifications for Wellington (Figure 9).

Nevertheless, cluster plots suggest that there may not be especially good separation between the clusters and that the clustering may be subject to some influence from outliers (Figure 11). This may be more of a problem for 5 clusters than for 4, where we see a smaller average silhouette width (Figure 12) and what appears to be almost complete overlap between cluster 4 and the other clusters when plotted over principal components 1 and 2 (Figure 11). This caveat should be considered when interpreting these results.

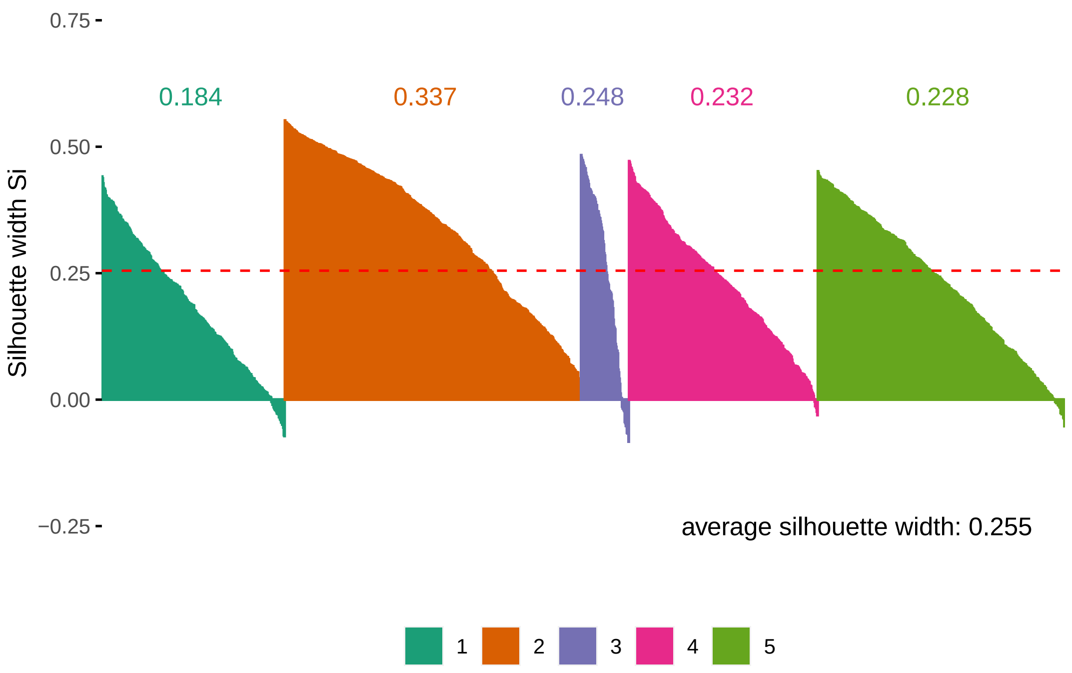
(a) 4 clusters

(b) 5 clusters

Figure 11: Cluster plots visualising the results of k-means clustering of daily weather data for Wellington, with 4 and 5 clusters. Dim 1 is principal component 1 and Dim2 is principal component 2.



(a) 4 clusterss



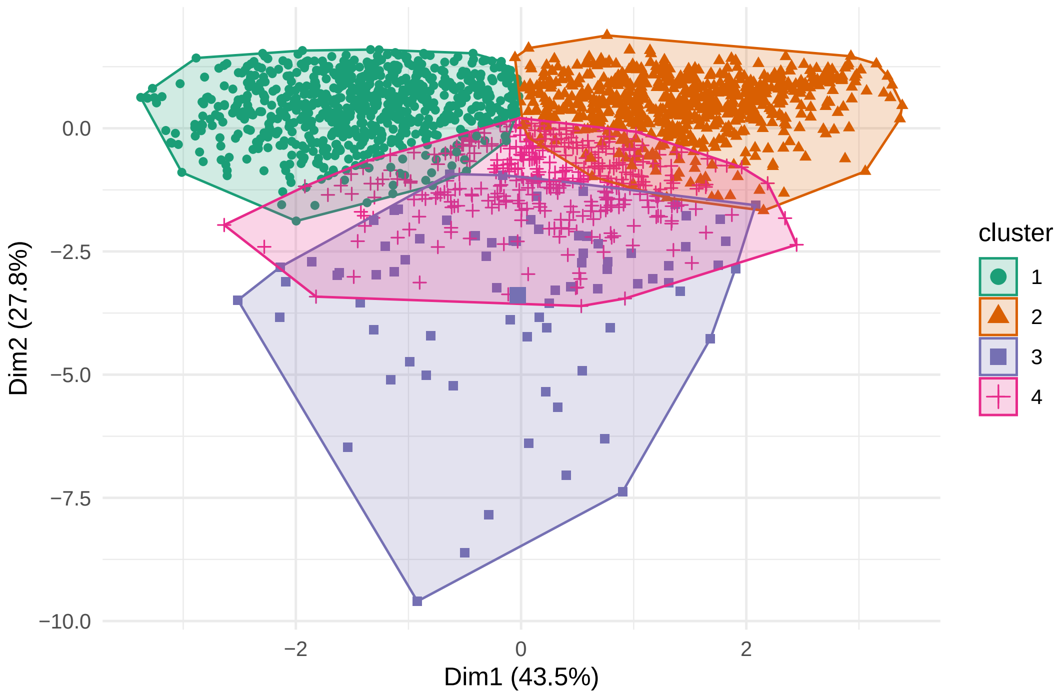
(b) 5 clusters

Figure 12: Silhouette plots of the k-means clusters of daily weather data for Wellington, comparing the results from specifying 4 and 5 clusters. The silhouette width for each cluster is displayed in the text above the cluster's values and the average silhouette width across all clusters is indicated by the red dashed line and the text at the bottom of the chart

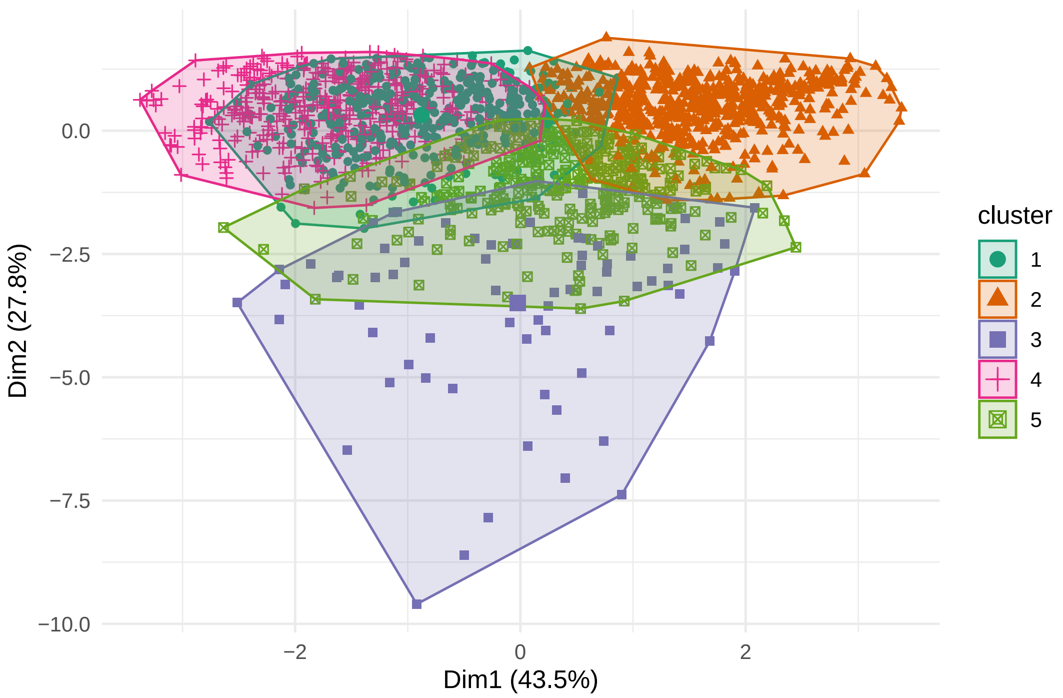
4.5.2 K-means clustering of Auckland daily weather data

The k-means clustering of standardised daily Auckland weather data also appears to generate clusters with greater in-group similarity than either of the two seasonal classifications. The silhouette widths for four and five k-means clusters (Figure 14) are consistently larger than the silhouette widths for the season and real season classifications for Auckland (Figure 10).

However, the clustering may be even more heavily influenced by outliers than in Wellington. Group 3 in both the 4- and the 5-cluster classifications has very few observations (Figure 14) and may consist predominantly of days that have unusually low values on principal component 2 (Figure 13). As for Wellington, the clusters do not appear to separate very well on principal components 1 and 2, especially for the 5-cluster scheme (Figure 13b). These issues should be considered when interpreting these results.

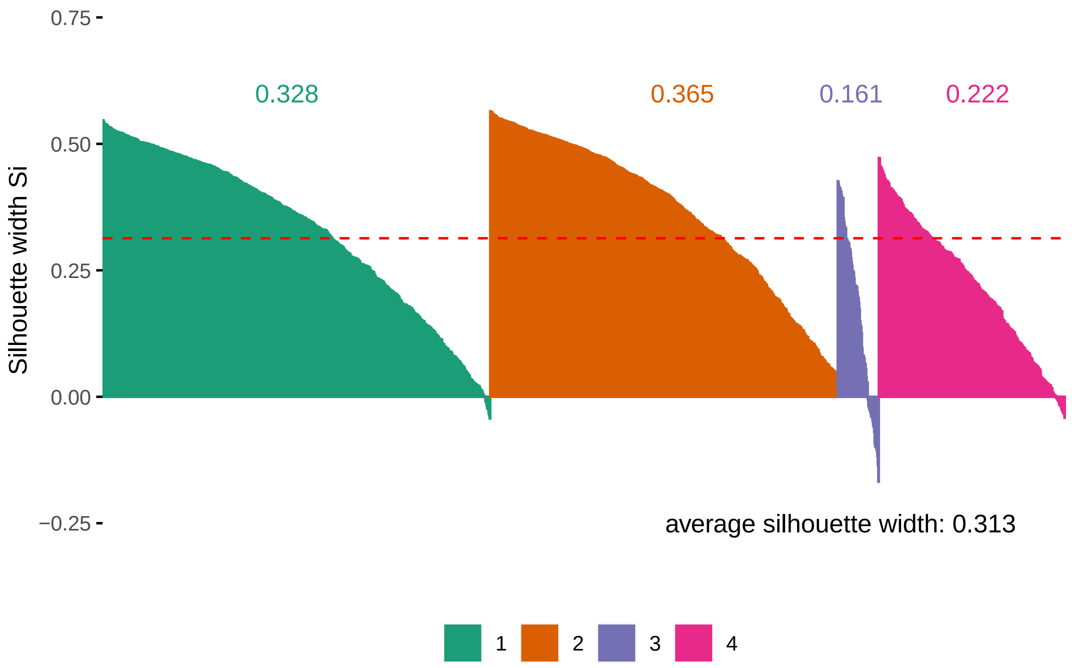


(a) 4 clusters

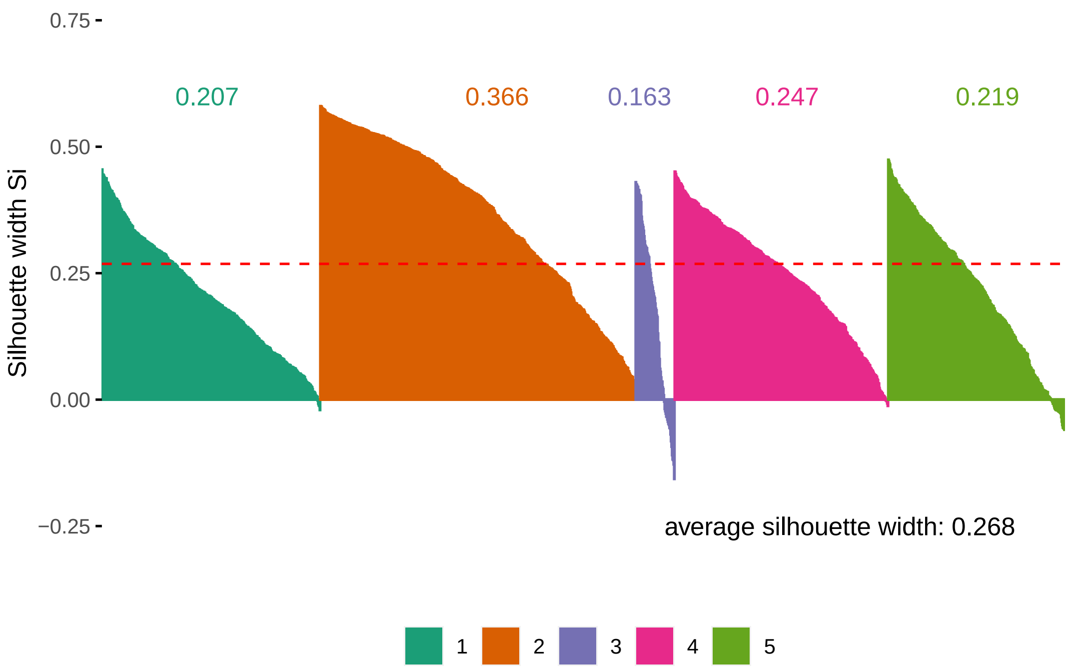


(b) 5 clusters

Figure 13: Cluster plots visualising the results of k-means clustering of daily weather data for Auckland, with 4 and 5 clusters. Dim 1 is principal component 1 and Dim2 is principal component 2.



(a) 4 clusters



(b) 5 clusters

Figure 14: Silhouette plots of the k-means clusters of daily weather data for Auckland, comparing the results from specifying 4 and 5 clusters. The silhouette width for each cluster is displayed in the text above the cluster's values and the average silhouette width across all clusters is indicated by the red dashed line and the text at the bottom of the chart

4.5.3 For both cities, the conventional seasons are more similar than the “real” seasons to k-means clusters generated from daily weather data

We show above that the 4- and 5-kmeans clusters perform better than season and real season at grouping together days with similar weather, for both cities. While there are caveats with the clustering results, the fact that they are consistently better than the seasonal classifications enables us to proceed with our comparison in which we use the clusters as a higher quality grouping of daily observations, against which we compare each seasonal classification.

The best match between each seasonal classification and the clusters was determined as described in the Methodology, and the confusion matrices for each of these best matches were generated (Tables 17, 18, 19, and 20). Seasons are given as row labels and clusters as column headings: cluster 1 was matched to the season in row 1, cluster 2 to the season in row 2, and so on.

Perfomance measures were calculated for each comparison as described in the Methodology. We find that, in both cities, the conventional season classification performs better than real season on every measure (Tables 21, 22, 23, and 24).

These results come with the caveat that they depend not just on the relative quality of the seasonal classifications, but also on the relative quality of the 4- versus 5-k-means clusters. Nevertheless, we obtain a result from this analysis that is consistent with our other lines of investigation. The comparison of k-means clusters with seasonal classifications provides additional evidence that the conventional seasons may be better than real seasons at classifying daily weather patterns.

Table 17: Confusion matrix comparing the conventional season classification with the four k-means clusters, Wellington

	1	2	3	4
Spring	157	24	133	119
Autumn	132	28	145	152
Winter	116	31	305	0
Summer	47	18	6	372

Table 18: Confusion matrix comparing the real season classification with the five k-means clusters, Wellington

	1	2	3	4	5
Shitsville	122	132	22	53	104
Winter	90	183	22	4	0
Spring1_2	55	98	16	41	94

	1	2	3	4	5
Autumn	64	133	22	176	62
Summer	8	3	7	76	198

Table 19: Confusion matrix comparing the conventional season classification with the four k-means clusters, Auckland

	1	2	3	4
Summer	413	0	17	21
Winter	0	330	24	106
Autumn	182	184	23	71
Spring	140	145	14	156

Table 20: Confusion matrix comparing the real season classification with the five k-means clusters, Auckland

	1	2	3	4	5
Autumn	210	146	21	24	59
Winter	0	228	19	0	58
Spring1_2	40	99	10	106	55
Summer	90	0	11	187	8
Shitsville	74	125	13	88	155

Table 21: Performance of the conventional season classification against the four k-means clusters, Wellington

Overall accuracy	0.483
Macro Precision	0.485
Macro Recall	0.430
Macro F1	0.432

Table 22: Performance of the real season classification against the five k-means clusters, Wellington

Overall accuracy	0.389
Macro Precision	0.402
Macro Recall	0.362
Macro F1	0.359

Table 23: Performance of the conventional season classification against the four k-means clusters, Auckland

Overall accuracy	0.505
Macro Precision	0.506
Macro Recall	0.450
Macro F1	0.439

Table 24: Performance of the real season classification against the five k-means clusters, Auckland

Overall accuracy	0.433
Macro Precision	0.442
Macro Recall	0.390
Macro F1	0.393

5 Conclusions

The key question for this investigation was "should we change the seasonal naming system for Wellington and/or Auckland". All of our analysis points towards the conventional season classification system being better than the suggested "real" season classification for both cities. This is clearly seen in the silhouette plots for both Auckland and Wellington, with the average silhouette width being higher for the conventional season classification than the real season classification. We also find that the conventional seasons have larger distances between seasonal summary statistics and lower standard deviations, suggesting that this classification does a better job of separating similar observations into more separate groups. And finally, we find that k-means clusters generated from weather data group days into categories that are more similar to the conventional season classification, with higher overall accuracy, macro precision, macro recall, and macro F1 scores for conventional seasons compared to 4 clusters than for real seasons compared to 5 clusters. We conclude that we do not have evidence to support a change to the seasonal naming system.

However, it is important to note that despite the conventional seasons performing better than the real seasons, neither system is particularly good at classifying the New Zealand weather. This matches the experiences we have had living in New Zealand, where it will pour with rain for days in the middle of summer, just as the family camping trip has begun. We also acknowledge that the primary determinant of the season is, in fact, the tilt of the Earth's axis relative to the sun. Weather is an incidental outcome of this tilt rather than a key driver of how seasons are classified.

There are several caveats with this investigation. Firstly, the comparison of seasons to k-means clusters is dependent not just on the quality of the seasonal classification but also on the quality of the k-means clusters. K-means clustering is one of many clustering methods and it is sensitive to outliers (which are present in our data). Further research could investigate the sensitivity of this result to different clustering methods (such as k-medoids and hierarchical clustering) and to different specifications of the method (such as requiring clusters to be copmprised of contiguous days).

Secondly, the Global Radiation variable, which measures total heat from the sun each day, is inherently tied to day length (sunlight hours), which is caused by the axial tilt of the Earth, and this may give conventional seasons an advantage over real seasons. Future research could redo the analysis without the Global Radiation variable and see if it changes the results.

Finally, our analysis only covers the five-year period 2017-2021 and there are some weather patterns that span multiple years such as the El Niño Southern Oscillation (NIWA 2022a). Global warming may also have affected seasonal weather patterns. It would be interesting to investigate whether the performance of the two seasonal classifications has changed over time, or whether it relates to different phases of the Southern Oscillation. It is possible that the 2014 real season proposal (Shand, A 2014) was preceded by a period of time during which the weather was a better fit with that classification. With regards to the overall poor performance of both classifications, it would be interesting to assess their performance over a longer time period, such as the last 50 years. Global warming may have made the weather less predictable, so perhaps conventional seasons were a more accurate classifier of New Zealand weather in the past.

6 References

Ministry for Culture and Heritage. 2021. *Laser Kiwi Flag*. https://nzhistory.govt.nz/media/photo/fire-lazar.

NIWA. 2022a. *El Niño and La Niña*. https://niwa.co.nz/climate/information-and-resources/elnino.

———. 2022b. *New Zealand's National Climate Database*. https://cliflo.niwa.co.nz/.

Shand, A. 2014. "For Those Who Are New to Wellington, or Those Who Forget Each Year, Here's a Realistic Guide to Wellington's Seasons." https://twitter.com/adamshand/status/5131970009 September 20, 2014.

Shand, A and Ivshin, D. 2022. *A Realistic New Zealand Calendar*. https://www.realnzweather.com/.

The Datsun Violets. 1996. *You Can't Beat Wellington (on a Good Day*.

The Mutton Birds. 1994. *Wellington*.

A Appendix A

A.1 Summary statistics for Wellington by season

Table 25: Summary statistics for spring weather in Wellington

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Sample size	455.000	455.000	433.000	455.000	455.000	227.000
Minimum	8.100	2.500	1.710	126.000	0.000	-1.609
First quartile	13.900	7.700	9.770	320.000	0.000	-0.367
Median	15.500	10.300	15.600	444.000	0.000	1.163
Third quartile	17.200	11.900	20.260	606.500	3.200	2.163
Maximum	23.700	15.600	30.420	1006.000	69.400	4.240
Mean	15.494	9.813	15.566	472.631	3.549	0.947
Standard deviation	2.750	2.816	7.105	185.025	7.920	1.595
Skewness	-0.103	-0.304	0.128	0.513	3.596	-0.145
Kurtosis	3.075	2.340	2.158	2.614	19.594	1.926

Table 26: Summary statistics for summer weather in Wellington

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Sample size	450.000	450.000	448.000	451.000	447.000	164.000
Minimum	12.400	5.800	2.560	135.000	0.000	-1.609
First quartile	18.400	12.300	15.675	332.500	0.000	-0.223
Median	20.400	14.400	22.030	435.000	0.000	1.163
Third quartile	22.400	16.000	27.242	563.500	1.200	2.219
Maximum	30.300	20.200	33.420	1093.000	73.400	4.296
Mean	20.476	14.099	20.903	460.452	2.883	0.966
Standard deviation	2.855	2.617	7.552	174.054	8.104	1.651
Skewness	0.136	-0.211	-0.468	0.748	4.609	-0.113
Kurtosis	2.868	2.592	2.242	3.355	29.530	1.961

Table 27: Summary statistics for autumn weather in Wellington

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Sample size	460.000	460.000	459.000	459.000	459.000	222.000
Minimum	7.900	3.400	0.960	111.000	0.000	-1.609
First quartile	15.000	9.475	6.335	288.500	0.000	-0.916

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Median	17.000	11.700	9.710	395.000	0.000	1.099
Third quartile	19.225	13.600	13.970	528.000	2.800	2.307
Maximum	26.500	18.500	25.070	993.000	68.200	4.222
Mean	17.079	11.465	10.402	423.948	3.700	0.820
Standard deviation	3.072	2.964	5.370	178.149	8.765	1.781
Skewness	0.017	-0.216	0.471	0.614	3.740	-0.089
Kurtosis	3.005	2.635	2.484	2.876	19.899	1.685

Table 28: Summary statistics for winter weather in Wellington

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Sample size	456.000	457.000	455.000	455.000	455.000	286.000
Minimum	6.700	0.900	0.360	82.000	0.000	-1.609
First quartile	11.500	5.700	3.920	268.500	0.000	-0.815
Median	12.800	7.900	6.070	397.000	0.400	1.030
Third quartile	14.025	9.400	8.035	560.500	4.800	2.208
Maximum	17.100	14.100	16.310	1168.000	59.400	4.084
Mean	12.675	7.729	6.145	423.508	4.476	0.846
Standard deviation	1.881	2.382	2.989	198.459	9.070	1.664
Skewness	-0.417	-0.077	0.416	0.590	3.208	-0.063
Kurtosis	3.004	2.428	2.876	3.014	14.863	1.835

A.2 Summary statistics for Auckland by season

Table 29: Summary statistics for spring weather in Auckland

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Sample size	455.000	455.000	455.000	455.000	455.000	209.000
Minimum	12.900	0.700	3.560	89.000	0.000	-1.609
First quartile	17.300	8.650	11.870	190.000	0.000	-0.223
Median	18.800	11.100	16.070	255.000	0.000	1.099
Third quartile	21.000	13.200	20.240	331.000	2.300	1.917
Maximum	26.800	19.200	30.900	628.000	53.000	3.970
Mean	19.105	10.886	16.489	267.042	2.600	0.855
Standard deviation	2.595	3.370	6.033	97.387	6.100	1.473

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Skewness	0.315	-0.229	0.262	0.669	4.360	-0.235
Kurtosis	2.680	2.648	2.427	3.186	27.451	2.096

Table 30: Summary statistics for summer weather in Auckland

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Sample size	451.000	451.000	451.000	451.000	451.000	134.000
Minimum	18.600	7.900	2.640	115.000	0.000	-1.609
First quartile	22.900	13.850	17.825	200.000	0.000	-0.815
Median	24.600	15.800	22.610	259.000	0.000	0.915
Third quartile	26.000	17.800	26.020	322.500	0.400	2.169
Maximum	29.800	22.100	33.720	614.000	52.000	3.951
Mean	24.435	15.694	21.563	268.896	2.147	0.859
Standard deviation	2.186	2.725	6.469	88.880	6.681	1.637
Skewness	-0.110	-0.176	-0.665	0.906	4.389	0.033
Kurtosis	2.462	2.431	3.011	4.195	24.207	1.863

Table 31: Summary statistics for autumn weather in Auckland

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Sample size	460.000	460.000	460.000	460.000	460.000	205.000
Minimum	12.600	0.400	1.590	0.000	0.000	-1.609
First quartile	19.500	9.500	8.297	141.750	0.000	-0.511
Median	21.350	12.300	10.965	188.500	0.000	1.030
Third quartile	23.500	14.925	15.192	261.000	2.050	2.054
Maximum	28.500	20.400	25.260	629.000	93.000	4.533
Mean	21.418	12.152	11.614	210.313	3.116	0.878
Standard deviation	2.797	3.912	4.855	97.354	8.919	1.557
Skewness	-0.151	-0.270	0.401	0.909	5.914	-0.005
Kurtosis	2.656	2.874	2.797	4.103	48.372	2.124

Table 32: Summary statistics for winter weather in Auckland

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Sample size	460.000	460.000	460.000	460.000	460.000	295.000

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Minimum	9.000	-1.400	1.160	60.000	0.000	-1.609
First quartile	15.100	4.900	5.850	144.750	0.000	-0.223
Median	16.150	7.300	7.840	195.000	0.600	1.099
Third quartile	17.100	9.825	9.565	278.500	5.200	2.197
Maximum	20.100	14.700	17.260	585.000	67.400	4.211
Mean	16.066	7.374	7.809	220.130	4.030	0.933
Standard deviation	1.630	3.346	2.908	104.547	7.305	1.519
Skewness	-0.478	-0.141	0.183	0.982	3.361	-0.239
Kurtosis	3.815	2.468	3.078	3.611	20.092	1.972

A.3 Summary statistics for Wellington by real season

Table 33: (#tab:summstatsWLG_rs_spring1_2)Summary statistics for spring1_2 weather in Wellington

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Sample size	309.000	308.000	305.000	308.000	307.000	151.000
Minimum	7.100	0.900	1.060	97.000	0.000	-1.609
First quartile	13.000	8.275	7.460	307.000	0.000	0.000
Median	15.400	10.600	10.970	432.500	0.000	1.163
Third quartile	19.500	13.525	22.860	564.000	3.100	2.342
Maximum	27.300	17.900	33.420	1093.000	73.400	4.296
Mean	16.240	10.638	14.565	449.542	3.964	1.050
Standard deviation	3.991	3.657	9.075	179.370	9.235	1.600
Skewness	0.334	-0.102	0.600	0.649	3.797	-0.140
Kurtosis	2.314	2.275	1.996	3.230	20.673	2.052

Table 34: (#tab:summstatsWLG_rs_shitsville)Summary statistics for shitsville weather in Wellington

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Sample size	455.000	455.000	433.000	455.000	455.000	227.000
Minimum	8.100	2.500	1.710	126.000	0.000	-1.609
First quartile	13.900	7.700	9.770	320.000	0.000	-0.367
Median	15.500	10.300	15.600	444.000	0.000	1.163
Third quartile	17.200	11.900	20.260	606.500	3.200	2.163
Maximum	23.700	15.600	30.420	1006.000	69.400	4.240

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Mean	15.494	9.813	15.566	472.631	3.549	0.947
Standard deviation	2.750	2.816	7.105	185.025	7.920	1.595
Skewness	-0.103	-0.304	0.128	0.513	3.596	-0.145
Kurtosis	3.075	2.340	2.158	2.614	19.594	1.926

Table 35: (#tab:summstatsWLG_rs_summer)Summary statistics for summer weather in Wellington

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Sample size	295.000	295.000	296.000	296.000	293.000	103.000
Minimum	14.400	6.500	2.560	135.000	0.000	-1.609
First quartile	19.000	12.600	16.277	327.000	0.000	-0.511
Median	20.900	14.700	21.690	431.500	0.000	0.875
Third quartile	23.000	16.600	26.495	553.750	0.800	2.175
Maximum	30.300	20.200	33.320	1025.000	62.000	4.127
Mean	20.981	14.463	20.758	454.986	2.195	0.761
Standard deviation	2.816	2.669	7.207	171.219	6.329	1.622
Skewness	0.092	-0.263	-0.519	0.688	5.003	-0.067
Kurtosis	2.736	2.546	2.428	3.200	36.461	1.889

Table 36: (#tab:summstatsWLG_rs_autumn)Summary statistics for autumn weather in Wellington

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Sample size	460.000	460.000	459.000	459.000	459.000	222.000
Minimum	7.900	3.400	0.960	111.000	0.000	-1.609
First quartile	15.000	9.475	6.335	288.500	0.000	-0.916
Median	17.000	11.700	9.710	395.000	0.000	1.099
Third quartile	19.225	13.600	13.970	528.000	2.800	2.307
Maximum	26.500	18.500	25.070	993.000	68.200	4.222
Mean	17.079	11.465	10.402	423.948	3.700	0.820
Standard deviation	3.072	2.964	5.370	178.149	8.765	1.781
Skewness	0.017	-0.216	0.471	0.614	3.740	-0.089
Kurtosis	3.005	2.635	2.484	2.876	19.899	1.685

Table 37: (#tab:summstatsWLG_rs_winter)Summary statistics for winter weather in Wellington

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Sample size	302.000	304.000	302.000	302.000	302.000	196.000
Minimum	6.700	1.600	0.360	82.000	0.000	-1.609
First quartile	11.400	5.700	3.500	255.500	0.000	-0.916
Median	12.700	7.700	5.065	388.500	0.400	1.064
Third quartile	13.900	9.300	7.198	562.750	4.950	2.225
Maximum	17.100	14.100	10.620	1168.000	59.400	4.084
Mean	12.539	7.675	5.212	421.275	4.852	0.834
Standard deviation	1.942	2.393	2.359	208.623	9.702	1.719
Skewness	-0.367	0.085	-0.018	0.630	3.080	-0.037
Kurtosis	2.898	2.454	2.109	3.025	13.685	1.758

A.4 Summary statistics for Auckland by real season

Table 38: (#tab:summstatsAKL_rs_spring1_2)Summary statistics for spring1_2 weather in Auckland

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Sample size	310.000	310.000	310.000	310.000	310.000	159.000
Minimum	9.000	-0.100	1.580	63.000	0.000	-1.609
First quartile	16.125	7.825	9.305	180.250	0.000	-0.511
Median	19.350	11.950	13.080	254.500	0.200	1.281
Third quartile	23.700	15.400	23.057	324.750	3.750	2.163
Maximum	27.900	20.700	32.890	575.000	52.000	3.951
Mean	19.809	11.557	15.942	260.484	3.230	0.921
Standard deviation	4.110	4.694	8.216	97.184	6.805	1.556
Skewness	0.073	-0.183	0.430	0.541	3.733	-0.304
Kurtosis	1.765	2.066	1.947	3.046	20.981	1.947

Table 39: (#tab:summstatsAKL_rs_shitsville)Summary statistics for shitsville weather in Auckland

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Sample size	455.000	455.000	455.000	455.000	455.000	209.000
Minimum	12.900	0.700	3.560	89.000	0.000	-1.609

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
First quartile	17.300	8.650	11.870	190.000	0.000	-0.223
Median	18.800	11.100	16.070	255.000	0.000	1.099
Third quartile	21.000	13.200	20.240	331.000	2.300	1.917
Maximum	26.800	19.200	30.900	628.000	53.000	3.970
Mean	19.105	10.886	16.489	267.042	2.600	0.855
Standard deviation	2.595	3.370	6.033	97.387	6.100	1.473
Skewness	0.315	-0.229	0.262	0.669	4.360	-0.235
Kurtosis	2.680	2.648	2.427	3.186	27.451	2.096

Table 40: (#tab:summstatsAKL_rs_summer)Summary statistics for summer weather in Auckland

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Sample size	296.000	296.000	296.000	296.000	296.000	81.000
Minimum	19.900	7.900	2.640	115.000	0.000	-1.609
First quartile	23.375	13.900	17.245	199.000	0.000	-0.916
Median	25.050	15.900	22.435	248.000	0.000	0.875
Third quartile	26.525	18.100	25.605	312.500	0.200	2.416
Maximum	29.800	22.100	33.720	614.000	43.200	3.766
Mean	24.928	15.866	21.193	263.980	2.084	0.883
Standard deviation	2.123	2.832	6.311	90.137	6.524	1.688
Skewness	-0.161	-0.193	-0.717	1.194	4.214	0.043
Kurtosis	2.460	2.361	3.101	5.035	22.183	1.680

Table 41: (#tab:summstatsAKL_rs_autumn)Summary statistics for autumn weather in Auckland

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Sample size	460.000	460.000	460.000	460.000	460.000	205.000
Minimum	12.600	0.400	1.590	0.000	0.000	-1.609
First quartile	19.500	9.500	8.297	141.750	0.000	-0.511
Median	21.350	12.300	10.965	188.500	0.000	1.030
Third quartile	23.500	14.925	15.192	261.000	2.050	2.054
Maximum	28.500	20.400	25.260	629.000	93.000	4.533
Mean	21.418	12.152	11.614	210.313	3.116	0.878
Standard deviation	2.797	3.912	4.855	97.354	8.919	1.557

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Skewness	-0.151	-0.270	0.401	0.909	5.914	-0.005
Kurtosis	2.656	2.874	2.797	4.103	48.372	2.124

Table 42: (#tab:summstatsAKL_rs_winter)Summary statistics for winter weather in Auckland

	Tmax	Tmin	GlobalRad	WindRun	Rainfall	logRainfall
Sample size	305.000	305.000	305.000	305.000	305.000	189.000
Minimum	11.000	-1.400	1.160	60.000	0.000	-1.609
First quartile	15.000	4.700	5.380	131.000	0.000	-0.223
Median	16.200	7.200	7.180	188.000	0.600	0.956
Third quartile	17.100	9.600	8.600	251.000	3.800	2.104
Maximum	20.100	14.700	13.100	585.000	67.400	4.211
Mean	16.035	7.184	6.891	208.669	3.948	0.912
Standard deviation	1.639	3.510	2.320	102.839	7.692	1.502
Skewness	-0.311	-0.095	-0.345	1.088	3.469	-0.106
Kurtosis	3.191	2.424	2.679	3.799	20.584	2.058