Data Mining Assignment 3: Bicycle Counter Data

Louis Carnec
15204934
September 1, 2016

In this assignment I will observe and analyse bicycle counter data for the Dún Laoghaire-Rathdown County Council for 2012-2015 using R. The excel file containing the data for the number of cyclists recorded at given locations will be pre-processed and handled to describe the data.
I will conduct this project by following a CRISP-DM framework (Cross Industry Standard Process for Data Mining). The phases of CRISP-DM are business understanding, data understanding, data preparation, modelling, evaluation and deployment. The first phase is to set out the project's objectives, in this case to uncover patterns in the data and create a predictive model for the number of cyclists at each location using the 2012-2014 data as a training dataset and using the 2015 data as a test set.
Next come the phases of data understanding and preparation.

1. Calculate monthly summaries at each location for the years 2012, 2013 and 2014.

(a) Describe the steps undertaken to achieve this task.

Data Understanding

Counter locations can be one-way or two-way. One way counters are represented by one variable, for example 'Rock.Road..Bus.Lane.Beside.Park'. There are three variables associated with each two-way counters; in numbers, out numbers and the sum of in (towards town) and out (away from town) numbers.
The exception is the counter for 'The Metals' at Glenageary dart station which takes 5 variables by recording both in and out for cyclists and pedestrians. The sum variable is the sum of both in and out cyclists and pedestrians, not just cyclists. Thus, if we were to compare the sum of in and out numbers for cyclists

across counters we would have to create a new variable for the sum of in and out cyclists at Glenageary dart station.
The four years' datasets have a different dimension, each of the four years have different numbers of variables (or different number of locations recorded).

2012 has 13 variables (excluding Date), 3 of which are two way counters (DLR Co. Co. Glenageary, N11 Montrose, Rock Road* Park), and two one-way counters (Rock Road* Bus Lane Beside Park, Totem N11 Stillorgan Rd.).

Listing 1: 2012 Column Names

```
> colnames(cycle_2012)
 [1] "Date"                              "DLR.Co..Co..
    Glenageary."
 [3] "Bicycle_IN"                        "Bicycle_OUT"
 [5] "Walking_IN"                        "Walking_OUT"
 [7] "N11.Montrose"                      "OUT.Away.From.
    City.Centre"
 [9] "IN..Towards.City.Centre"           "Rock.Road..Bus.
    Lane.Beside.Park"
[11] "Rock.Road..Park"                   "
   Rock_Road_Park_IN"
[13] "Rock_Road_Park_OUT"                "Totem.N11.
   Stillorgan.Rd."
[15] "Month"
```

2013 has 17 variables (excluding Date). A new two-way location is added to the dataset, Totem Clonskeagh Road (making up three variables; sum of in and out, in, out) and a one-way location, Totem Rock Road.

Listing 2: 2013 Column Names

```
> colnames(cycle_2013)
 [1] "Date"                              "DLR.Co..Co..
    Glenageary."
 [3] "Bicycle_IN"                        "Bicycle_OUT"
 [5] "Walking_IN"                        "Walking_OUT"
 [7] "N11.Montrose"                      "OUT.Away.From.
    City.Centre"
 [9] "IN..Towards.City.Centre"           "Rock.Road..Bus.
    Lane.Beside.Park"
```

```
[11] "Rock.Road..Park"                    "
   Rock_Road_Park_IN"
[13] "Rock_Road_Park_OUT"                 "Totem.Clonskeagh
   .Road"
[15] "Totem.IN"                           "Totem.OUT"
[17] "Totem.N11.Stillorgan.Rd."           "Totem.Rock.Road"
[19] "Month"
```

2014 has 16 variables (excluding Date). The one-way location 'Totem N11 Stillorgan Rd.' was removed from the 2013 locations.

Listing 3: 2014 Column Names

```
> colnames(cycle_2014)
 [1] "Date"                               "DLR.Co..Co..
    Glenageary."
 [3] "Bicycle_IN"                         "Bicycle_OUT"
 [5] "Walking_IN"                         "Walking_OUT"
 [7] "N11.Montrose"                       "OUT.Away.From.
    City.Centre"
 [9] "IN..Towards.City.Centre"            "Rock.Road..Bus.
    Lane.Beside.Park"
[11] "Rock.Road..Park"                    "
   Rock_Road_Park_IN"
[13] "Rock_Road_Park_OUT"                 "Totem.Clonskeagh
   .Road"
[15] "Totem.IN"                           "Totem.OUT"
[17] "Totem.Rock.Road"                    "Month"
```

2015 has 19 locations (excluding Date). The same locations as 2013 can be found in the 2015 dataset plus two new locations 'N11 ECO TOTEM' and 'Rock Road Inbound' are added. Also, two of the variables were renamed from 'Totem Clonskeagh Road' to 'Totem Clonskeagh Road. Data Only' and 'Stillorgan Rock Road' to 'Stillorgan Rock Road. Data Only', these two locations have no observations however.

Listing 4: 2014 Column Names

```
> colnames(cycle_2015)
 [1] "Date"                                   "Glenageary."
 [3] "Walking_IN"                             "Walking_OUT"
 [5] "Bicycle_IN"                             "Bicycle_OUT"
```

```
 [7] "N11.ECO.TOTEM"                          "N11.Montrose
   "
 [9] "OUT.Away.From.City.Centre"              "IN..Towards.
   City.Centre"
[11] "Rock.Road..Bus.Lane.Beside.Park"        "Rock.Road..
  Park"
[13] "Rock_Road_Park_IN"                      "
  Rock_Road_Park_OUT"
[15] "Rock.Road.Inbound"                      "Totem.
  Clonskeagh.Road..Data.Only."
[17] "Totem.IN"                               "Totem.OUT"
[19] "Totem.N11.Stillorgan.Rd..Data.Only." "Totem.Rock.
  Road..Data.Only."
[21] "Month"
```

Data Preparation

The data for each location starts at different dates due to the fact that the totems were not put in place at the same time, meaning that there is missing or unobserved data for some extended periods of time. Further, there are short periods of time with missing data due to machine malfunction or battery requiring replacement. We must take account of the missing data. In R, the blank spaces were replaced by NAs.

Listing 5: Setting Blank Spaces to NA

```
for (i in (2:14)){
  cycle_2012[,i] <- as.character(cycle_2012[,i])
  cycle_2012[,i] <- as.numeric(cycle_2012[,i])
  cycle_2012[,i][is.na(cycle_2012[,i])]<-NA}
```

Some locations, such as Rock Road Park have values of zero (November and December 2014 for example). This is likely to be an error, therefore we count these instances as unobserved data and replace the zeros with NAs.

Listing 6: Setting zeros to NA

```
cycle_2012[cycle_2012==0]<-NA
```

Missing values can impact the reliability of summary statistics. In calculating monthly statistics we will ignore these missing values. This will affect the monthly sum of locations for which values are missing, however other 'averaged' statistics will not take account or ignore missing values. This may however bias or skew the monthly statistics. For example if values are missing for the first 22/23 days of January (which is the case in the 2012 dataset) the mean for January could be underestimated. We could imagine that more people cycle to work at the start of January as a new year's resolution and that numbers would start falling as January goes on and the weather deteriorates.

We can look for missing values by running through each column using a for loop.

Listing 7: Looking for missing values

```
for (i in (1:15)){
  print(colnames(cycle_2012)[i])
  print(is.na(cycle_2012[,i]))
}
```

'N11 Montrose' and the 'Rock Road' locations in the first 22/23 days of the month of January 2012 have missing values (NAs).
In 2013, the 'Clonskeagh Road' location has three missing values for the sum of in and out and for in and out, these can be ignored as they would have a small impact on summary statistics.
The 2014 dataset has missing values from the beginning of November to the end of December for the 'N11 Montrose' location and the 'Rock Road' location. The Clonskeagh road has missing data from the beginning of March to the beginning of September.
The 2015 dataset has missing values for 'N11 ECO totem' for the whole year except from the end of Decembre. The 'N11 Montrose' and the 'Rock Road Park' locations have missing values for the first three months of 2015. 'Rock Road Inbound' has missing values for the first 10 months of 2015. And lastly, 'Totem N11 Stillorgan Rd. Data Only' and 'Totem Rock Road. Data Only' have no values.

In producing monthly statistics, dates have to be pre-processed to extract the month associated with each date. To do so I used the 'lubridate' package.

Listing 8: Extracting Months

```
cycle_2012$Month <- month(ymd(cycle_2012$Date))
cycle_2013$Month <- month(ymd(cycle_2013$Date))
cycle_2014$Month <- month(ymd(cycle_2014$Date))
cycle_2015$Month <- month(ymd(cycle_2015$Date))
```

We may need to bind the three years' datasets in the second part of this assignment. We would then need columns for the year of each row (see listing 8).

Listing 9: Extracting Years

```
cycle_2012$Year <- year(ymd(cycle_2012$Date))
cycle_2013$Year <- year(ymd(cycle_2013$Date))
cycle_2014$Year <- year(ymd(cycle_2014$Date))
cycle_2015$Year <- year(ymd(cycle_2015$Date))
```

In calculating monthly summaries for 'The Metals' location, we create a new variable for the sum of the number of cyclists going in and out (see listing 10).

Listing 10: Extracting Years

```
cycle_2012$Glen.Bicycle <- cycle_2012$Bicycle_IN +
   cycle_2012$Bicycle_OUT
cycle_2013$Glen.Bicycle <- cycle_2013$Bicycle_IN +
   cycle_2013$Bicycle_OUT
cycle_2014$Glen.Bicycle <- cycle_2014$Bicycle_IN +
   cycle_2014$Bicycle_OUT
cycle_2015$Glen.Bicycle <- cycle_2015$Bicycle_IN +
   cycle_2015$Bicycle_OUT
```

This variable will allow us to compare 'The metals' to other two-way locations directly.

Lastly we bind the three training datasets into one. This will allow us to create monthly statistics on a three year basis and we will have a complete three year training dataset. One of obstacles to building a single dataset for the three years comes from the fact that all years do not have the same number of locations. A solution is to pick the year with the least number of locations (where other years include all these locations) and retain only those locations for every other year.
2012 has the fewest locations. We remove all the locations not included in 2012 from 2013 and 2014. We can then bind the three data sets (see listing 11). Alternatively, we can add new locations to 2012 and 2013 and fill them with NA columns for those locations which they do not observe.

Listing 11: Training dataset 1: Removing columns and binding

```
#Creating a single data set
cycle_2012_tobind <- cycle_2012[-14]

cycle_2013_tobind <- cycle_2013
cycle_2013_tobind <- cycle_2013_tobind[-14] # *3
cycle_2013_tobind <- cycle_2013_tobind[-15]
cycle_2013_tobind <- cycle_2013_tobind[-15]
cycle_2013_tobind <- cycle_2013_tobind[-14]
```

```
dim(cycle_2012_tobind)
dim(cycle_2013_tobind)
dim(cycle_2014_tobind)

cycle_2014_tobind <- cycle_2014_tobind[-14] # *3
cycle_2014_tobind <- cycle_2014[-15]

library(dplyr)
cycle_training <- rbind(cycle_2012_tobind,
   cycle_2013_tobind)
cycle_training <- rbind(cycle_training,cycle_2014_tobind
   )
```

Alternatively, we pursue a second training dataset where we add NA columns to have the same columns for 2012,2013 and 2014. In this case, we have 21 columns for each year. Four NA columns were added to 2012 for the Clonskeagh Location and the Rock Road location, the Stillorgan Road location was added to 2012. We then bound all three years into our second training data set (see listing).

Listing 12: Training dataset 2: Adding columns and binding

```
cycle_2012$Totem.Clonskeagh.Road <- NA
cycle_2012$Totem.IN <- NA
cycle_2012$Totem.OUT <- NA
cycle_2012$Totem.Rock.Road <- NA

cycle_2014$Totem.N11.Stillorgan.Rd. <-NA

dim(cycle_2012)
dim(cycle_2013)
dim(cycle_2014)

cycle_training2 <- rbind(cycle_2012,cycle_2013)
cycle_training2 <- rbind(cycle_training2,cycle_2014)
```

Data Understanding (bis)

After pre-processing the data we can calculate monthly summaries at each location for the years 2012, 2013 and 2014. The monthly summaries will allow us to better understand the data and trends in the data.

Tabular Summaries

In listing 12 the monthly summaries for the training dataset (2012-2014) are calculated, these include the monthly sum, mean, median, standard deviation, minimum and maximum.

Listing 13: Calculating Monthly Summaries

```
aggregate(data.frame(cycle_training[,15],cycle_training
   [,3:4],cycle_training[,7:13]),by=list(
   cycle_training$Month), FUN=sum ,na.rm=TRUE)
aggregate(data.frame(cycle_training[,15],cycle_training
   [,3:4],cycle_training[,7:13]),by=list(
   cycle_training$Month), mean,na.rm=TRUE)
aggregate(data.frame(cycle_training[,15],cycle_training
   [,3:4],cycle_training[,7:13]),by=list(
   cycle_training$Month), median,na.rm=TRUE)
aggregate(data.frame(cycle_training[,15],cycle_training
   [,3:4],cycle_training[,7:13]),by=list(
   cycle_training$Month), sd,na.rm=TRUE)
aggregate(data.frame(cycle_training[,15],cycle_training
   [,3:4],cycle_training[,7:13]),by=list(
   cycle_training$Month), min,na.rm=TRUE)
aggregate(data.frame(cycle_training[,15],cycle_training
   [,3:4],cycle_training[,7:13]),by=list(
   cycle_training$Month), max,na.rm=TRUE)
```

(b) Present the results in both tabular form and as histograms.

Monthly summaries for all three years

Month	The Metals	The Metals_IN	The Metals_OUT	N11.Montrose	Montrose OUT	Montrose IN	Rock Road Bus Lane Beside Park	Rock Road Park	Rock_Road_Park_IN	Rock_Road_Park_OUT
1	10887.00	5467.00	5420.00	33193.00	27744.00	5449.00	24934.00	18890.00	8682.00	10208.00
2	11149.00	6042.00	5107.00	43335.00	35977.00	7358.00	35062.00	23615.00	10918.00	12697.00
3	14658.00	8361.00	6297.00	45273.00	38096.00	7177.00	41708.00	28048.00	12840.00	15208.00
4	16321.00	9060.00	7261.00	49325.00	41923.00	7402.00	44881.00	31332.00	14475.00	16857.00
5	18169.00	9623.00	8546.00	49299.00	42580.00	6719.00	55475.00	39052.00	18047.00	21005.00
6	20492.00	10826.00	9666.00	43505.00	37824.00	5681.00	53187.00	43328.00	20185.00	23143.00
7	22101.00	11828.00	10273.00	47804.00	41769.00	6035.00	59356.00	52075.00	23930.00	28145.00
8	21852.00	11668.00	10184.00	42975.00	37472.00	5503.00	50832.00	41666.00	19089.00	22577.00
9	21220.00	11155.00	10065.00	66714.00	57128.00	9586.00	56052.00	47508.00	21580.00	25928.00
10	16287.00	8433.00	7854.00	66318.00	56405.00	9913.00	49667.00	36696.00	16420.00	20276.00
11	13105.00	6496.00	6609.00	43834.00	36689.00	7145.00	41309.00	18486.00	8477.00	10009.00
12	10229.00	5055.00	5174.00	20787.00	17489.00	3298.00	27214.00	12713.00	5793.00	6920.00

Table 1: 2012-2014 Monthly SUM

Month	The Metals	The Metals_IN	The Metals_OUT	N11.Montrose	Montrose OUT	Montrose IN	Rock Road Bus Lane Beside Park	Rock Road Park	Rock_Road_Park_IN	Rock_Road_Park_OUT
1	117.06	58.78	58.28	481.06	402.09	78.97	361.36	269.86	124.03	145.83
2	131.16	71.08	60.08	509.82	423.26	86.56	412.49	277.82	128.45	149.38
3	157.61	89.90	67.71	486.81	409.63	77.17	448.47	301.59	138.06	163.53
4	181.34	100.67	80.68	548.06	465.81	82.24	498.68	348.13	160.83	187.30
5	195.37	103.47	91.89	530.10	457.85	72.25	596.51	419.91	194.05	225.86
6	227.69	120.29	107.40	483.39	420.27	63.12	590.97	481.42	224.28	257.14
7	237.65	127.18	110.46	514.02	449.13	64.89	638.24	559.95	257.31	302.63
8	234.97	125.46	109.51	462.10	402.92	59.17	546.58	448.02	205.26	242.76
9	235.78	123.94	111.83	741.27	634.76	106.51	622.80	527.87	239.78	288.09
10	175.13	90.68	84.45	713.10	606.51	106.59	534.05	398.87	178.48	220.39
11	145.61	72.18	73.43	644.62	539.54	105.07	458.99	308.10	141.28	166.82
12	109.99	54.35	55.63	329.95	277.60	52.35	292.62	205.05	93.44	111.61

Table 2: 2012-2014 Monthly Mean

Month	The Metals	The Metals_IN	The Metals_OUT	N11.Montrose	Montrose OUT	Montrose IN	Rock Road Bus Lane Beside Park	Rock Road Park	Rock_Road_Park_IN	Rock_Road_Park_OUT
1	123.00	61.00	61.00	561.00	477.00	85.00	411.00	257.00	125.00	135.00
2	134.00	67.00	61.00	590.00	483.00	102.00	431.00	266.00	126.00	146.00
3	162.00	89.00	63.00	519.00	427.00	77.00	443.00	284.00	135.00	147.00
4	178.00	93.00	79.00	622.50	523.00	91.50	529.00	302.00	149.50	151.00
5	192.00	99.00	94.00	598.00	512.00	77.00	615.00	386.00	186.00	206.00
6	223.00	120.50	107.00	537.00	465.50	67.50	589.00	441.00	205.00	240.50
7	241.00	124.00	111.00	576.00	501.00	71.00	660.00	507.00	236.00	275.00
8	242.00	126.00	112.00	534.00	452.00	61.00	568.00	434.00	205.00	230.00
9	236.00	121.50	111.50	811.50	696.50	114.50	689.50	506.00	233.50	271.00
10	170.00	89.00	83.00	790.00	660.00	115.00	581.00	357.00	169.00	194.50
11	144.50	74.00	74.00	740.00	617.00	126.50	514.50	283.00	132.00	143.50
12	113.00	55.00	56.00	216.00	178.00	42.00	238.00	172.00	81.50	84.50

Table 3: 2012-2014 Monthly Median

Month	The Metals	The Metals_IN	The Metals_OUT	N11.Montrose	Montrose OUT	Montrose IN	Rock Road Bus Lane Beside Park	Rock Road Park	Rock_Road_Park_IN	Rock_Road_Park_OUT
1	37.15	19.30	19.02	256.97	217.90	42.20	161.15	111.39	47.74	65.43
2	46.07	28.16	22.55	249.61	213.04	39.63	174.14	108.10	48.93	61.08
3	58.44	35.85	29.01	265.08	226.74	41.34	209.59	146.82	63.85	84.27
4	60.89	36.80	28.29	269.10	235.78	36.44	175.76	181.01	74.72	107.87
5	62.06	33.32	30.10	220.19	197.85	24.45	197.17	184.52	80.23	106.45
6	79.45	44.32	39.38	235.47	212.18	25.53	224.54	261.83	116.03	147.14
7	63.10	34.14	31.56	208.99	189.09	21.37	194.68	261.94	114.96	148.60
8	64.46	34.32	32.02	207.38	186.82	22.51	178.01	167.16	72.64	96.99
9	74.83	38.55	38.25	356.06	313.82	47.21	206.81	222.15	92.99	130.51
10	52.56	27.31	27.25	340.07	298.85	44.57	188.82	173.93	70.75	104.64
11	35.17	18.50	18.41	319.37	273.34	47.51	173.58	124.78	54.33	71.51
12	34.63	17.76	18.20	246.79	211.88	35.56	167.66	132.80	57.27	76.52

Table 4: 2012-2014 Monthly Standard Deviation

2012 Monthly Summaries

	Group.1	The Metals	The Metals_IN	The Metals_OUT	N11.Montrose	Montrose OUT	Montrose IN	Rock Road Bus Lane Beside Park	Rock Road Park	Rock_Road_Park_IN	Rock_Road_Park_OUT	Totem N11 Stillorgan Rd
1	1.00	3680.00	1900.00	1780.00	3184.00	2579.00	605.00	2834.00	1311.00	638.00	673.00	11120.00
2	2.00	4268.00	2245.00	2023.00	14728.00	12049.00	2679.00	15001.00	6872.00	3392.00	3480.00	13255.00
3	3.00	5684.00	3041.00	2643.00	16525.00	13633.00	2892.00	18914.00	10002.00	4743.00	5259.00	14141.00
4	4.00	4442.00	2354.00	2088.00	13192.00	10973.00	2219.00	14467.00	6790.00	3303.00	3487.00	12056.00
5	5.00	5456.00	2923.00	2533.00	15033.00	12886.00	2147.00	19551.00	10791.00	5133.00	5658.00	13895.00
6	6.00	5073.00	2654.00	2419.00	11397.00	9852.00	1545.00	15753.00	8371.00	4079.00	4292.00	10618.00
7	7.00	5829.00	3110.00	2719.00	12692.00	10972.00	1720.00	17619.00	9222.00	4506.00	4716.00	11285.00
8	8.00	6764.00	3703.00	3061.00	12724.00	11133.00	1591.00	17216.00	10204.00	5066.00	5138.00	11411.00
9	9.00	6303.00	3395.00	2908.00	17638.00	15064.00	2574.00	17065.00	9933.00	4869.00	5064.00	15367.00
10	10.00	5138.00	2658.00	2480.00	18774.00	16001.00	2773.00	15879.00	8241.00	3981.00	4260.00	16549.00
11	11.00	4201.00	2094.00	2107.00	17503.00	14419.00	3084.00	13325.00	6754.00	3259.00	3495.00	9685.00
12	12.00	3138.00	1545.00	1593.00	9158.00	7626.00	1532.00	8263.00	4549.00	2225.00	2324.00	6631.00

Table 5: 2012 Monthly SUM

	Group.1	The Metals	The Metals_IN	The Metals_OUT	N11.Montrose	Montrose OUT	Montrose IN	Rock Road Bus Lane Beside Park	Rock Road Park	Rock_Road_Park_IN	Rock_Road_Park_OUT	Totem N11 Stillorgan Rd
1	1.00	118.71	61.29	57.42	454.86	368.43	86.43	404.86	163.88	79.75	84.12	358.71
2	2.00	147.17	77.41	69.76	507.86	415.48	92.38	517.28	236.97	116.97	120.00	457.07
3	3.00	183.35	98.10	85.26	533.06	439.77	93.29	610.13	322.65	153.00	169.65	456.16
4	4.00	148.07	78.47	69.60	439.73	365.77	73.97	482.23	226.33	110.10	116.23	401.87
5	5.00	176.00	94.29	81.71	484.94	415.68	69.26	630.68	348.10	165.58	182.52	448.23
6	6.00	169.10	88.47	80.63	379.90	328.40	51.50	525.10	279.03	135.97	143.07	353.93
7	7.00	188.03	100.32	87.71	409.42	353.94	55.48	568.35	297.48	145.35	152.13	364.03
8	8.00	218.19	119.45	98.74	410.45	359.13	51.32	555.35	329.16	163.42	165.74	368.10
9	9.00	210.10	113.17	96.93	587.93	502.13	85.80	568.83	331.10	162.30	168.80	512.23
10	10.00	165.74	85.74	80.00	605.61	516.16	89.45	512.23	265.84	128.42	137.42	533.84
11	11.00	140.03	69.80	70.23	583.43	480.63	102.80	444.17	225.13	108.63	116.50	322.83
12	12.00	101.23	49.84	51.39	295.42	246.00	49.42	266.55	146.74	71.77	74.97	213.90

Table 6: 2012 Monthly Mean

	Group.1	The Metals	The Metals_IN	The Metals_OUT	N11.Montrose	Montrose OUT	Montrose IN	Rock Road Bus Lane Beside Park	Rock Road Park	Rock_Road_Park_IN	Rock_Road_Park_OUT	Totem N11 Stillorgan Rd
1	1.00	126.00	65.00	58.00	561.00	443.00	108.00	472.00	182.00	86.00	86.50	441.00
2	2.00	134.00	68.00	67.00	606.00	492.00	105.00	571.00	235.00	119.00	115.00	546.00
3	3.00	170.00	97.00	77.00	600.00	499.00	96.00	673.00	297.00	139.00	159.00	541.00
4	4.00	150.50	78.00	72.00	480.00	404.00	79.50	517.50	247.50	116.00	124.00	446.50
5	5.00	178.00	88.00	82.00	525.00	455.00	74.00	633.00	291.00	144.00	156.00	489.00
6	6.00	177.50	92.50	83.50	366.50	315.50	51.00	524.50	275.50	138.50	133.50	331.50
7	7.00	197.00	105.00	94.00	477.00	410.00	62.00	627.00	312.00	150.00	160.00	431.00
8	8.00	212.00	112.00	99.00	441.00	395.00	55.00	554.00	356.00	174.00	169.00	411.00
9	9.00	204.50	103.00	92.50	677.50	582.50	89.50	595.00	330.00	161.50	171.00	586.50
10	10.00	162.00	84.00	76.00	712.00	595.00	99.00	581.00	274.00	134.00	142.00	619.00
11	11.00	139.00	72.50	71.50	657.50	542.50	120.00	488.50	246.00	119.00	122.00	228.00
12	12.00	109.00	53.00	53.00	215.00	176.00	42.00	214.00	146.00	64.00	73.00	125.00

Table 7: 2012 Monthly Median

	Group.1	The Metals	The Metals_IN	The Metals_OUT	N11.Montrose	Montrose OUT	Montrose IN	Rock Road Bus Lane Beside Park	Rock Road Park	Rock_Road_Park_IN	Rock_Road_Park_OUT	Totem N11 Stillorgan Rd
1	1.00	31.19	17.23	15.42	246.88	203.19	44.42	173.34	75.32	38.61	37.73	212.37
2	2.00	49.20	28.65	23.42	227.20	186.79	41.42	185.01	68.47	33.02	36.27	204.32
3	3.00	56.11	25.04	32.96	253.50	209.38	45.82	182.10	113.01	54.55	59.54	233.83
4	4.00	49.09	23.95	26.37	213.11	181.67	32.54	166.48	73.39	36.22	38.15	197.80
5	5.00	69.11	38.64	31.51	201.89	181.83	22.19	213.46	159.72	79.13	81.62	190.78
6	6.00	69.73	37.96	32.79	196.06	175.46	22.80	212.53	139.35	64.39	75.51	182.36
7	7.00	53.12	26.62	27.92	165.45	145.75	20.50	166.59	99.07	47.50	53.27	150.32
8	8.00	70.46	36.10	35.83	168.67	154.57	16.38	170.13	113.12	56.32	57.77	153.72
9	9.00	84.98	44.66	41.20	271.30	238.13	38.22	207.68	113.77	57.54	57.22	246.76
10	10.00	52.31	27.81	26.56	272.82	240.08	38.03	178.31	77.67	37.89	41.30	244.68
11	11.00	33.36	15.90	18.57	273.48	228.50	45.66	175.61	65.42	33.29	32.97	247.16
12	12.00	26.16	14.09	13.93	190.53	161.13	29.86	139.34	56.35	30.55	26.71	148.22

Table 8: 2012 Monthly Standard Deviation

2013 Monthly Summaries

	Group.1	The Metals	The Metals_IN	The Metals_OUT	N11 Montrose	Montrose OUT	Montrose IN	Rock Road Bus Lane Beside Park	Rock Road Park	Rock_Road_Park_IN	Rock_Road_Park_OUT	Totem Clonskeagh Road	Clon. IN	Clon. OUT	Totem N11 Stillorgan Rd	Totem Rock Road
1	1.00	3443.00	1710.00	1733.00	15932.00	13261.00	2671.00	10693.00	8006.00	3920.00	4086.00	34937.00	17956.00	16981.00	9944.00	10800.00
2	2.00	3210.00	1652.00	1558.00	15187.00	12619.00	2568.00	10277.00	8248.00	3860.00	4388.00	43268.00	22086.00	21182.00	12615.00	11609.00
3	3.00	3381.00	1758.00	1623.00	11761.00	9917.00	1844.00	9122.00	5443.00	2621.00	2822.00	33937.00	17383.00	16554.00	9949.00	9811.00
4	4.00	4943.00	2698.00	2245.00	16703.00	14074.00	2629.00	14560.00	8648.00	4306.00	4342.00	42944.00	21964.00	20980.00	13626.00	12664.00
5	5.00	6072.00	3284.00	2788.00	16624.00	14289.00	2335.00	18376.00	11031.00	5370.00	5661.00	46475.00	23942.00	22533.00	14356.00	16949.00
6	6.00	7190.00	3996.00	3194.00	15235.00	13177.00	2058.00	18049.00	13571.00	6567.00	7004.00	38172.00	19532.00	18640.00	10291.00	15947.00
7	7.00	8262.00	4581.00	3681.00	17727.00	15455.00	2272.00	22075.00	20959.00	9833.00	11126.00	43567.00	22051.00	21516.00	11545.00	19581.00
8	8.00	8136.00	4378.00	3758.00	15807.00	13711.00	2096.00	18145.00	15575.00	7168.00	8407.00	39364.00	20107.00	19257.00	11197.00	16147.00
9	9.00	6892.00	3731.00	3161.00	21933.00	18642.00	3291.00	18401.00	15632.00	7108.00	8524.00	57647.00	29548.00	28099.00	3832.00	16455.00
10	10.00	5269.00	2811.00	2458.00	22127.00	18666.00	3461.00	16279.00	12583.00	5624.00	6959.00	61514.00	31375.00	30139.00	0.00	14275.00
11	11.00	4606.00	2280.00	2326.00	20631.00	17397.00	3234.00	14409.00	11732.00	5218.00	6514.00	55857.00	27931.00	27926.00	0.00	13725.00
12	12.00	3659.00	1776.00	1883.00	11578.00	9818.00	1760.00	9294.00	8164.00	3568.00	4596.00	32686.00	17027.00	15659.00	0.00	8628.00

Table 9: 2013 Monthly SUM

	Group.1	The Metals	The Metals_IN	The Metals_OUT	N11 Montrose	Montrose OUT	Montrose IN	Rock Road Bus Lane Beside Park	Rock Road Park	Rock_Road_Park_IN	Rock_Road_Park_OUT	Totem Clonskeagh Road	Clon. IN	Clon. OUT	Totem N11 Stillorgan Rd	Totem Rock Road
1	1.00	111.06	55.16	55.90	513.94	427.77	86.16	344.94	258.26	126.45	131.81	1127.00	579.23	547.77	320.77	348.39
2	2.00	114.64	59.00	55.64	542.39	450.68	91.71	367.04	294.57	137.86	156.71	1545.29	788.79	756.50	450.54	414.61
3	3.00	109.06	56.71	52.35	379.39	319.90	59.48	294.26	175.58	84.55	91.03	1094.74	560.74	534.00	320.94	316.48
4	4.00	164.77	89.93	74.83	556.77	469.13	87.63	485.33	288.27	143.53	144.73	1590.52	813.48	777.04	486.64	469.04
5	5.00	195.87	105.94	89.94	536.26	460.94	75.32	592.77	355.84	173.23	182.61	1499.19	772.32	726.87	463.10	546.74
6	6.00	239.67	133.20	106.47	507.83	439.23	68.60	601.63	452.37	218.90	233.47	1272.40	651.07	621.33	343.03	531.57
7	7.00	266.52	147.77	118.74	571.84	498.55	73.29	712.10	676.10	317.19	358.90	1405.39	711.32	694.06	372.42	631.65
8	8.00	262.45	141.23	121.23	509.90	442.29	67.61	585.32	502.42	231.23	271.19	1269.81	648.61	621.19	361.19	520.87
9	9.00	229.73	124.37	105.37	731.10	621.40	109.70	613.37	521.07	236.93	284.13	1921.57	984.93	936.63	479.00	548.50
10	10.00	169.97	90.68	79.29	713.77	602.13	111.65	525.13	405.90	181.42	224.48	1984.32	1012.10	972.23		460.48
11	11.00	153.53	76.00	77.53	687.70	579.90	107.80	480.30	391.07	173.93	217.13	1861.90	931.03	930.87		457.50
12	12.00	118.03	57.29	60.74	373.48	316.71	56.77	299.81	263.35	115.10	148.26	1054.39	549.26	505.13		278.32

Table 10: 2013 Monthly Mean

	Group.1	The Metals	The Metals_IN	The Metals_OUT	N11 Montrose	Montrose OUT	Montrose IN	Rock Road Bus Lane Beside Park	Rock Road Park	Rock_Road_Park_IN	Rock_Road_Park_OUT	Totem Clonskeagh Road	Clon. IN	Clon. OUT	Totem N11 Stillorgan Rd	Totem Rock Road
1	1.00	116.00	56.00	60.00	606.00	492.00	102.00	385.00	252.00	121.00	133.00	1332.00	686.00	644.00	383.00	399.00
2	2.00	120.50	61.50	54.50	621.00	496.00	111.00	389.00	295.00	136.00	159.00	1746.50	889.00	852.00	509.00	443.50
3	3.00	94.00	50.00	45.00	384.00	326.00	58.00	273.00	154.00	75.00	79.00	1253.00	637.00	601.00	351.00	303.00
4	4.00	165.00	87.50	77.00	640.00	534.00	98.50	496.00	286.50	147.50	147.00	1825.00	940.00	909.00	556.50	498.00
5	5.00	187.00	101.00	85.00	566.00	487.00	84.00	602.00	377.00	177.00	188.00	1694.00	876.00	820.00	490.00	548.00
6	6.00	245.50	125.00	104.00	537.00	465.50	70.00	586.50	401.50	192.50	220.50	1352.50	691.50	677.00	279.00	513.50
7	7.00	271.00	156.00	119.00	643.00	561.00	76.00	710.00	668.00	311.00	355.00	1562.00	800.00	773.00	431.00	628.00
8	8.00	274.00	143.00	125.00	581.00	499.00	80.00	648.00	517.00	218.00	285.00	1416.00	741.00	689.00	345.00	585.00
9	9.00	236.50	124.50	111.00	848.50	727.00	119.00	655.50	513.00	232.00	273.50	2248.00	1160.00	1085.50	583.50	577.50
10	10.00	172.00	90.00	80.00	792.00	660.00	122.00	551.00	420.00	186.00	236.00	2285.00	1150.00	1120.00		451.00
11	11.00	149.50	75.50	76.50	832.50	699.00	130.50	547.00	420.00	181.00	233.00	2216.50	1148.50	1151.50		515.50
12	12.00	124.00	59.00	64.00	230.00	191.00	44.00	238.00	200.00	88.00	112.00	686.00	364.00	322.00		170.00

Table 11: 2013 Monthly Median

	Group.1	The Metals	The Metals_IN	The Metals_OUT	N11 Montrose	Montrose OUT	Montrose IN	Rock Road Bus Lane Beside Park	Rock Road Park	Rock_Road_Park_IN	Rock_Road_Park_OUT	Totem Clonskeagh Road	Clon. IN	Clon. OUT	Totem N11 Stillorgan Rd	Totem Rock Road
1	1.00	39.20	20.59	19.70	251.20	213.04	42.29	146.07	59.53	27.79	32.81	559.74	292.14	268.12	164.19	161.39
2	2.00	33.33	17.32	17.55	277.33	241.01	40.54	128.76	103.62	50.25	55.58	705.35	360.37	345.27	220.25	164.50
3	3.00	46.52	25.22	22.42	231.63	197.25	35.51	139.74	79.68	37.95	42.79	631.56	325.70	306.15	198.72	171.73
4	4.00	50.00	28.69	23.79	254.12	215.10	41.82	177.96	90.01	46.13	45.95	656.05	336.57	319.98	207.30	175.42
5	5.00	61.97	32.59	30.73	209.96	185.75	26.09	176.73	115.96	57.46	60.24	562.21	290.00	272.75	186.81	184.33
6	6.00	69.82	43.93	33.18	233.51	208.33	26.93	230.39	178.98	85.86	94.13	601.79	307.35	294.90	184.39	214.17
7	7.00	48.66	29.60	23.35	200.81	180.46	21.71	190.45	194.10	93.29	101.85	518.16	265.42	253.63	145.26	183.96
8	8.00	51.60	29.57	24.46	225.19	200.45	26.20	186.75	140.15	64.24	77.12	547.23	281.62	266.08	156.46	184.99
9	9.00	64.90	34.38	33.11	333.85	292.40	44.25	207.51	178.32	78.68	100.32	915.97	464.52	451.90	238.08	196.25
10	10.00	57.79	30.79	29.24	339.72	296.37	45.64	201.62	143.98	65.58	78.93	924.17	469.29	455.39		188.87
11	11.00	24.58	12.39	13.55	319.25	272.57	47.37	170.26	114.70	51.88	63.86	879.72	462.98	434.46		192.17
12	12.00	45.55	22.69	23.85	289.46	249.80	40.25	195.70	160.19	69.04	91.75	799.12	411.62	388.08		219.77

Table 12: 2013 Monthly Standard Deviation

2014 Monthly Summaries

	Group.1	The Metals	The Metals_IN	The Metals_OUT	N11 Montrose	Montrose OUT	Montrose IN	Rock Road Bus Lane Beside Park	Rock Road Park	Rock_Road_Park_IN	Rock_Road_Park_OUT	Totem Clonskeagh Road	Clon. IN	Clon. OUT	Totem Rock Road
1	1.00	3764.00	1857.00	1907.00	14077.00	11904.00	2173.00	11407.00	9573.00	4124.00	5449.00	39508.00	20568.00	18940.00	11415.00
2	2.00	3671.00	2145.00	1526.00	13420.00	11309.00	2111.00	9784.00	8495.00	3666.00	4829.00	30807.00	21306.00	9501.00	9878.00
3	3.00	5593.00	3562.00	2031.00	16987.00	14546.00	2441.00	13672.00	12603.00	5476.00	7127.00	7427.00	6817.00	610.00	13833.00
4	4.00	6936.00	4008.00	2928.00	19430.00	16876.00	2554.00	15854.00	15894.00	6866.00	9028.00	0.00	0.00	0.00	14388.00
5	5.00	6641.00	3416.00	3225.00	17642.00	15405.00	2237.00	17548.00	17230.00	7544.00	9686.00	0.00	0.00	0.00	16050.00
6	6.00	8229.00	4176.00	4053.00	16873.00	14795.00	2078.00	19385.00	21386.00	9539.00	11847.00	0.00	0.00	0.00	16926.00
7	7.00	8010.00	4137.00	3873.00	17385.00	15342.00	2043.00	19662.00	21894.00	9591.00	12303.00	2200.00	2192.00	8.00	18461.00
8	8.00	6952.00	3587.00	3365.00	14444.00	12628.00	1816.00	15471.00	15887.00	6855.00	9032.00	0.00	0.00	0.00	14428.00
9	9.00	8025.00	4029.00	3996.00	27143.00	23422.00	3721.00	20586.00	21943.00	9603.00	12340.00	59909.00	28851.00	31058.00	8581.00
10	10.00	5880.00	2964.00	2916.00	25417.00	21738.00	3679.00	17509.00	15872.00	6815.00	9057.00	63132.00	30853.00	32279.00	0.00
11	11.00	4298.00	2122.00	2176.00	5700.00	4873.00	827.00	13575.00	0.00	0.00	0.00	53425.00	26002.00	27423.00	0.00
12	12.00	3432.00	1734.00	1698.00	51.00	45.00	6.00	9657.00	0.00	0.00	0.00	32834.00	16169.00	16665.00	0.00

Table 13: 2014 Monthly SUM

	Group.1	The Metals	The Metals_IN	The Metals_OUT	N11 Montrose	Montrose OUT	Montrose IN	Rock Road Bus Lane Beside Park	Rock Road Park	Rock_Road_Park_IN	Rock_Road_Park_OUT	Totem Clonskeagh Road	Clon. IN	Clon. OUT	Totem Rock Road
1	1.00	121.42	59.90	61.52	454.10	384.00	70.10	367.97	308.81	133.03	175.77	1274.45	663.48	610.97	368.23
2	2.00	131.11	76.61	54.50	479.29	403.89	75.39	349.43	303.39	130.93	172.46	1100.25	760.93	339.32	352.79
3	3.00	180.42	114.90	65.52	547.97	469.23	78.74	441.03	406.55	176.65	229.90	928.38	852.12	76.25	446.23
4	4.00	231.20	133.60	97.60	647.67	562.53	85.13	528.47	529.80	228.87	300.93				496.14
5	5.00	214.23	110.19	104.03	569.10	496.94	72.16	566.06	555.81	243.35	312.45				517.74
6	6.00	274.30	139.20	135.10	562.43	493.17	69.27	646.17	712.87	317.97	394.90				564.20
7	7.00	258.39	133.45	124.94	560.81	494.90	65.90	634.26	706.26	309.39	396.87	733.33	730.67	8.00	595.52
8	8.00	224.26	115.71	108.55	465.94	407.35	58.58	499.06	512.48	221.13	291.35				465.42
9	9.00	267.50	134.30	133.20	904.77	780.73	124.03	686.20	731.43	320.10	411.33	2139.61	1030.39	1109.21	660.08
10	10.00	189.68	95.61	94.06	819.90	701.23	118.68	564.81	529.07	227.17	301.90	2036.52	995.26	1041.26	
11	11.00	143.27	70.73	72.53	712.50	609.12	103.38	452.50				1780.83	866.73	914.10	
12	12.00	110.71	55.94	54.77	51.00	45.00	6.00	311.52				1059.16	521.58	537.58	

Table 14: 2014 Monthly Mean

	Group.1	The Metals	The Metals_IN	The Metals_OUT	N11 Montrose	Montrose OUT	Montrose IN	Rock Road Bus Lane Beside Park	Rock Road Park	Rock_Road_Park_IN	Rock_Road_Park_OUT	Totem Clonskeagh Road	Clon. IN	Clon. OUT	Totem Rock Road
1	1.00	126.00	61.00	64.00	546.00	470.00	69.00	425.00	339.00	155.00	194.00	1423.00	740.00	683.00	440.00
2	2.00	140.50	72.00	52.50	533.50	456.50	82.00	385.50	298.50	130.50	170.50	1122.50	866.00	190.50	356.00
3	3.00	175.00	112.00	62.00	649.00	568.00	83.00	463.00	430.00	178.00	252.00	1132.50	1102.00	31.00	485.00
4	4.00	235.50	137.00	101.00	755.00	652.00	96.00	582.50	547.50	238.00	309.50				548.00
5	5.00	215.00	109.00	104.00	654.00	583.00	78.00	630.00	568.00	246.00	322.00				549.00
6	6.00	278.50	139.00	135.50	637.50	569.00	75.00	681.50	631.00	284.50	352.50				609.50
7	7.00	258.00	133.00	126.00	638.00	565.00	71.00	681.00	688.00	297.00	386.00	918.00	918.00	8.00	635.00
8	8.00	224.00	115.00	108.00	534.00	452.00	70.00	540.00	528.00	220.00	315.00				502.00
9	9.00	250.00	123.00	125.00	969.00	846.00	134.00	779.50	755.50	325.50	432.00	2585.50	1252.50	1333.00	749.00
10	10.00	192.00	96.00	97.00	999.00	868.00	132.00	611.00	550.50	231.50	315.50	2498.00	1235.00	1217.00	
11	11.00	143.50	65.50	76.00	864.50	735.50	129.00	501.00				2210.50	1071.00	1126.50	
12	12.00	113.00	56.00	55.00	51.00	45.00	6.00	284.00				788.00	376.00	412.00	

Table 15: 2014 Monthly Median

	Group.1	The Metals	The Metals_IN	The Metals_OUT	N11 Montrose	Montrose OUT	Montrose IN	Rock Road Bus Lane Beside Park	Rock Road Park	Rock_Road_Park_IN	Rock_Road_Park_OUT	Totem Clonskeagh Road	Clon. IN	Clon. OUT	Totem Rock Road
1	1.00	40.77	20.01	21.58	269.00	229.56	41.30	175.37	137.73	59.25	79.33	779.74	405.00	374.85	194.48
2	2.00	49.23	33.02	23.49	247.64	213.79	35.66	155.57	134.15	59.61	74.94	626.61	395.20	355.35	173.94
3	3.00	38.75	28.73	20.62	282.07	248.97	35.88	174.33	136.88	57.95	79.88	523.27	476.92	97.77	191.97
4	4.00	49.67	31.76	27.14	299.60	267.02	34.05	184.48	184.04	77.02	108.09				193.98
5	5.00	49.40	26.62	24.05	245.21	221.49	25.34	200.62	191.85	79.89	113.30				200.39
6	6.00	60.74	32.69	32.36	242.97	222.20	23.35	220.74	245.51	112.33	134.93				189.95
7	7.00	56.33	27.67	30.12	221.85	204.41	18.55	203.73	235.95	100.54	136.56	367.61	372.22		207.75
8	8.00	62.67	32.29	31.61	218.18	198.38	21.48	171.44	176.94	78.22	100.80				182.46
9	9.00	63.16	33.97	30.84	389.81	345.87	51.47	194.44	154.08	64.37	92.09	1022.67	507.03	521.58	174.57
10	10.00	45.37	22.69	23.95	375.62	332.71	45.61	187.98	175.31	68.05	109.09	959.12	469.74	490.36	
11	11.00	44.45	24.91	21.97	463.56	404.56	60.14	178.59				891.99	435.09	457.71	
12	12.00	27.76	14.90	14.39				165.42				711.76	353.44	358.79	

Table 16: 2014 Monthly Standard Deviation

2012-2014 Histograms

Listing 14: Generating Histograms

```
for (i in (1:15)){
  pdf(paste("trainingmean",i,".pdf",sep=""))
  hist(trainingmean[,i],breaks=5,xlab="Mean Monthly
     Number Cyclist Count",main=names(cycle_training
     [2:16])[i])
  dev.off()}

for (i in (1:20)){
  pdf(paste("training2mean",i,".pdf",sep=""))
  hist(training2mean[,i],breaks=5,xlab="Mean Monthly
     Number Cyclist Count",main=names(cycle_training2
     [2:21])[i])
  dev.off()}
```

The Metals: The two-way histogram for the metals is skewed to the left. This can be explained by the fact that cyclists are more likely to use the Metals in the sunnier months of the year. People that may not normally commute on the bike in the rainier months of the year would use the path in the sunnier months (accounting for a high count 4/12 months of the year). Further the path is often flooded during the winter which may account for a lower count in the winter.

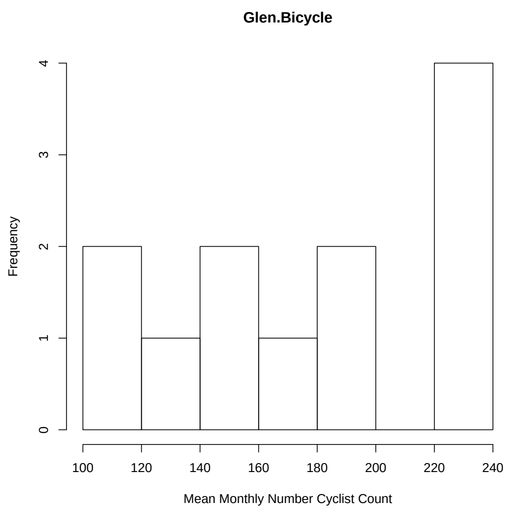

Figure 1: The Metals two-way (Bicycle only)

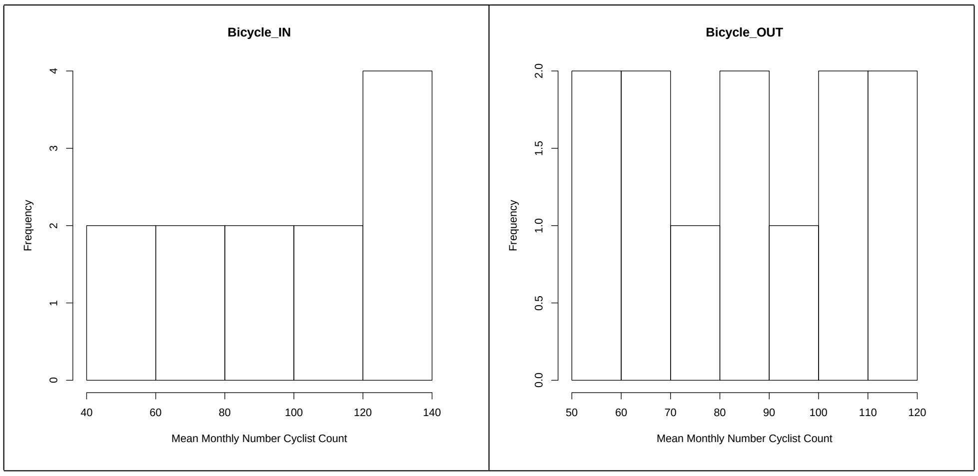

Figure 2: The Metals In and Out

N11 Montrose: The majority of months have on average 400 to 600 cyclists passing by the counter each day.

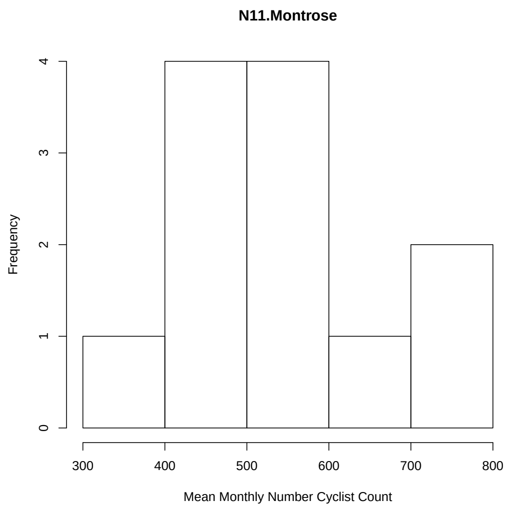

Figure 3: N11 Montrose two-way

There are a lot more cyclists passing by the away from town counter than past the in towards the city centre counter. The count difference between the in and out data can be explained by two variables, the position of counter and the state of the cycle lane. Where the counter is positioned on the way into town, the cycle path is elevated and its surface is both inconsistent (up and down) and badly surfaced, therefore some cyclists may avoid taking it, preferring riding in the bus lane instead. The cycling lane coming home from town is better surfaced and on an uphill incline, it is safer as cyclists are taking it at lower speeds and therefore cyclist count is higher. The distribution of the counter for cyclists coming out of town is also more consistent, out of 12 months of the year, 8 months of the year have between 400 and 500 cyclists passing by the counter.

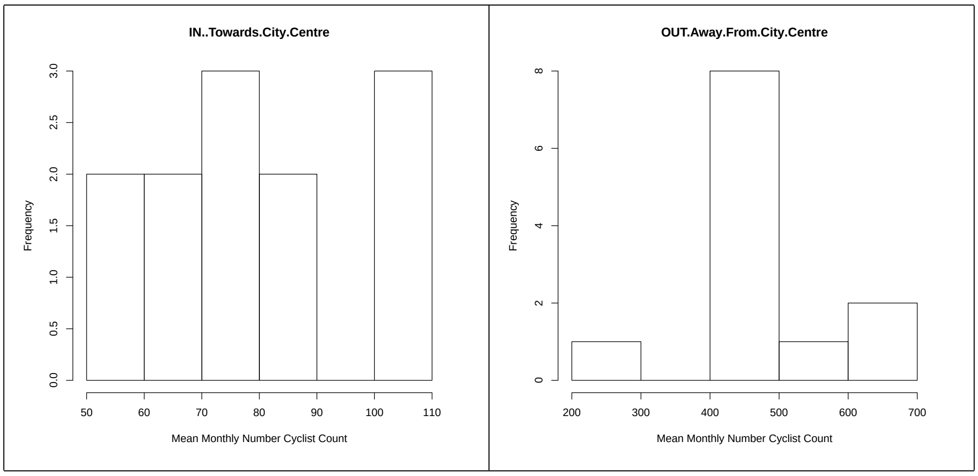

Figure 4: N11 Montrose In and Out

Totem Rock Road: The data on Totem Rock Road is limited (starts in Jan 2013 and finished Sept 2014). The cyclist count seems to be normally distributed.
Rock Road Bus Lane Beside Park: This cycling lane is positioned on the side of the road in the direction away form town. It is therefore taken by those commuters coming home from town that do not go through the park. The cyclist count is uniformly distributed between 400 and 650 cyclists.

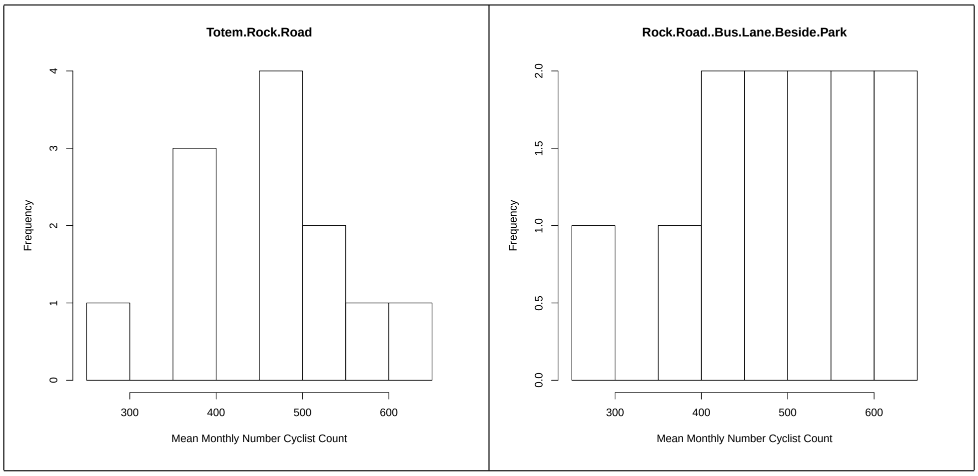

Figure 5: Totem Rock Road and Rock Road Bus Lane Beside Park

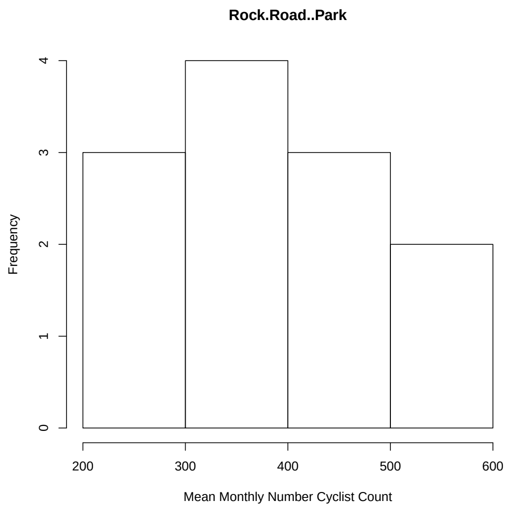

Figure 6: Rock Road Park two-way

Rock Road Park: More cyclists go through Blackrock park, more consistently, on the way back from town than on the way in. Commuters tend to be in more of a rush in the morning, as a result they may prefer to commute towards town in the morning by the Rock Road rather than through Blackrock Park.

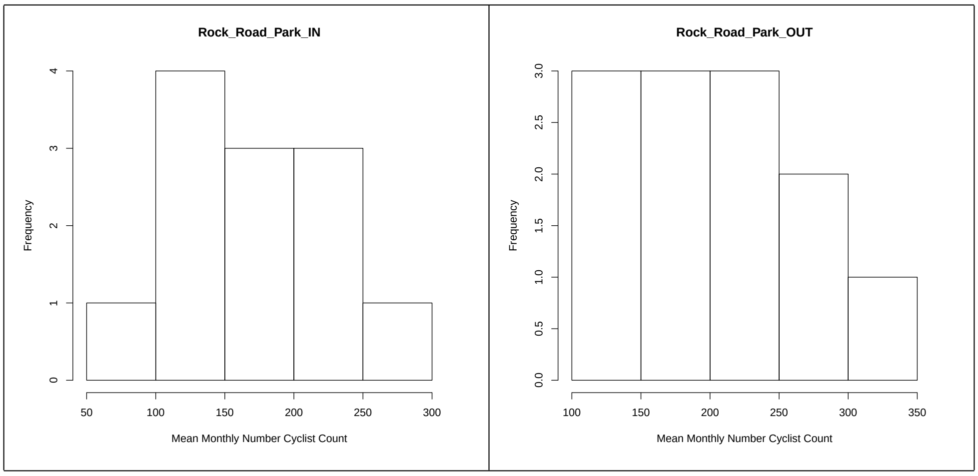

Figure 7: DTW between devices

Clonskeagh Road: The cyclist count for Clonskeagh is high compared to other counter locations (between 1000 and 2200 cyclists for the two-way count. The counts and distributions of the towards town and back from town data are roughly the same.

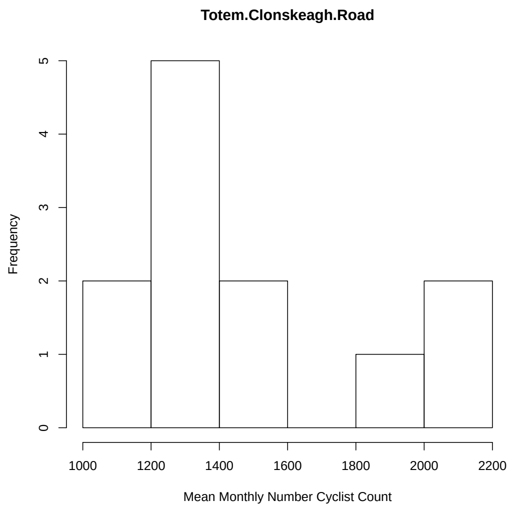

Figure 8: Clonskeagh Road

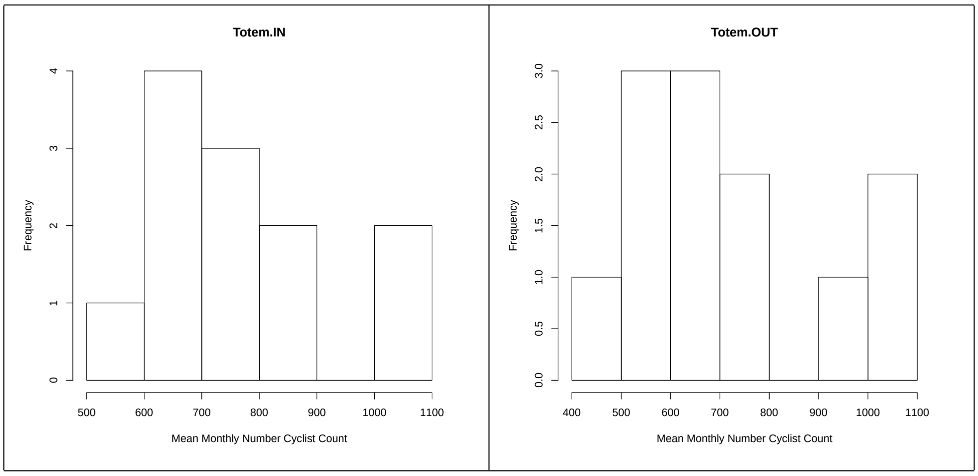

Figure 9: Clonskeagh Road In and Out

N11 Stillorgan Road: The N11 Stillorgan Road location is roughly normally distributed.

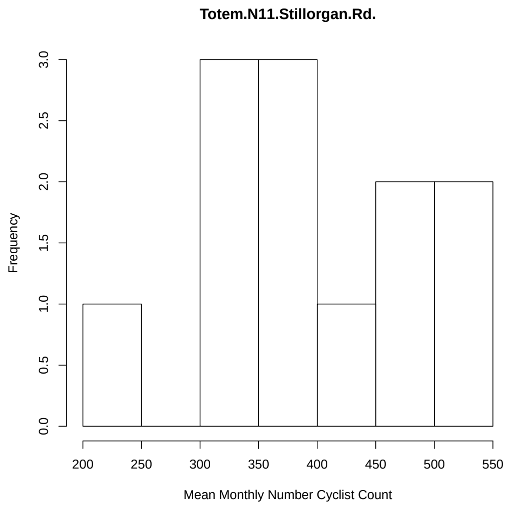

Figure 10: N11 Stillorgan Road

(c) Describe any trends observed in the monthly data (d) Attempt to explain any trends observed.

Average Mean Monthly Count Time Series

Listing 15: Generating Average Mean Monthly Count Time Series

```
trainingmean <-apply(cycle_training[,2:16], 2, function(
   x) tapply(x, cycle_training$Month, mean,na.rm=TRUE))
training2mean <-apply(cycle_training2[,2:21], 2,
   function(x) tapply(x, cycle_training$Month, mean,na.
   rm=TRUE))


for (i in (1:15)){
  pdf(paste("trainingmeants", i,".pdf",sep=""))
  plot(trainingmean[,i],ylab='Number of cyclists',main=
     names(cycle_training[2:16])[i],xaxt='n', lty=3)
  axis(1, at=1:12, labels=c('January','February','March
     ','April','May','June','July','August','Septembre
     ','October','November','December'),las=2)
  lines(trainingmean[,i], type="o", pch=22, lty=2, col="
     blue")
  dev.off()}

for (i in (1:20)){
  pdf(paste("training2meants", i,".pdf",sep=""))
  plot(training2mean[,i],ylab='Number of cyclists',main=
     names(cycle_training2[2:21])[i],xaxt='n', lty=3)
  axis(1, at=1:12, labels=c('January','February','March
     ','April','May','June','July','August','Septembre
     ','October','November','December'),las=2)
  lines(training2mean[,i], type="o", pch=22, lty=2, col
     ="blue")
  dev.off()}
```

The metals (see Figure 11 and 12): There is a bell shaped trend in the cyclist count, where the highest count can be found in the summer months (June, July, August, September). As temperatures get warmer and weather gets generally better, more people use cycling as a mode of transport and thus cyclist count increases.
Unlike the rock road locations (see Figure 13, 14, 15) the number cyclists does not fall in august. We could hypothesise that the metals' cyclist demographic is made up of a higher proportion of recreational cyclists versus commuters, therefore the fall in cyclist count is not as large during the month of August when a large proportion of commuters are on holidays.

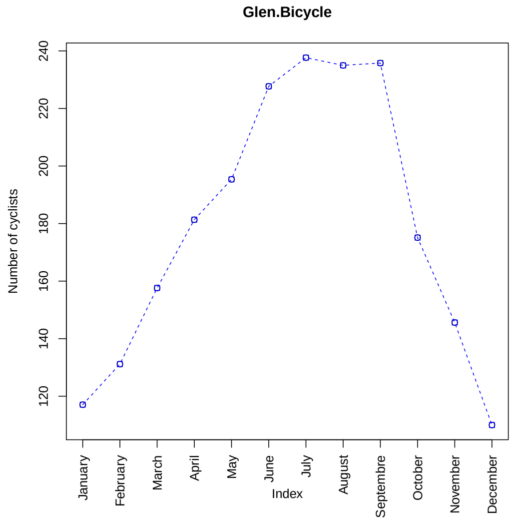

Figure 11: The Metals two-way (Bicycle only)

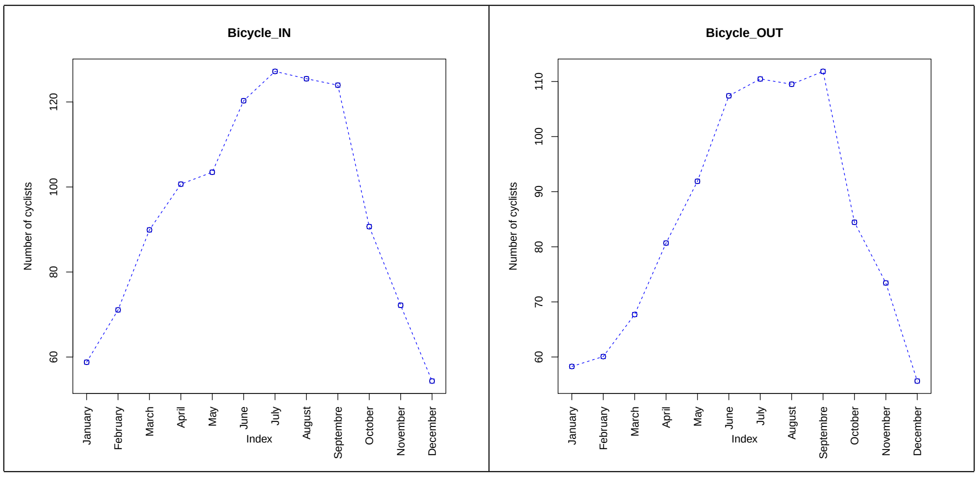

Figure 12: The Metals In and Out

The Rock Road (see Figure 13, 14 and 15): The distribution of cyclist count is similar to that of the metals with a higher count in the summer months. As we hypothesised, the fall in August could be due to a lower number of commuters as people may be away on holidays.

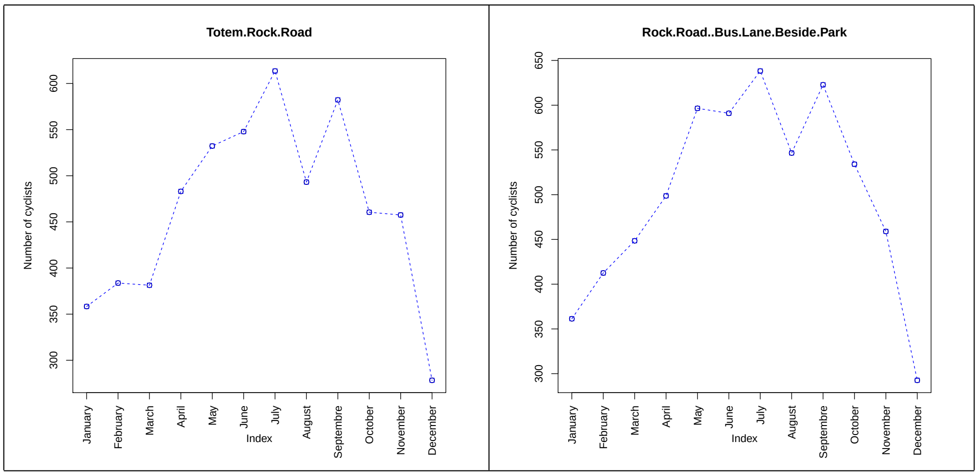

Figure 13: Totem Rock Road and Rock Road Bus Lane Beside Park

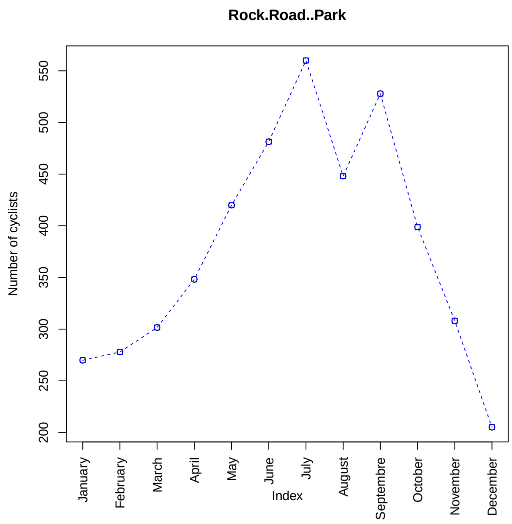

Figure 14: Rock Road Park two-way

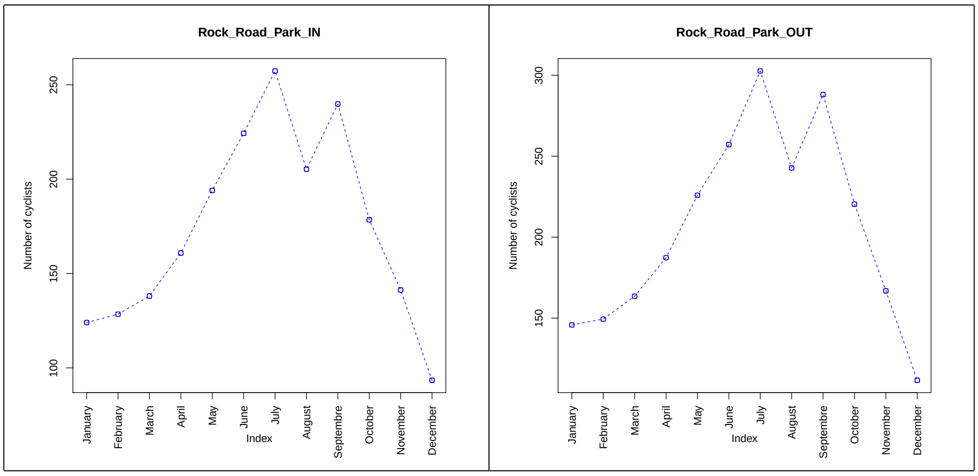

Figure 15: Rock Road Park In and Out

N11 Montrose (see Figure 16 and 17):
The N11 Montrose two-way count is stable around the 500 mark between January and August, this number jumps to above 700 in September and October and falls back down to the lowest count of the year in December (see Figure 16). The fact that UCD is located beside the N11 Montrose counters may explain why the count jumps upwards in September and falls dramatically in December. The semester starts in September for UCD students and ends in December for three weeks.

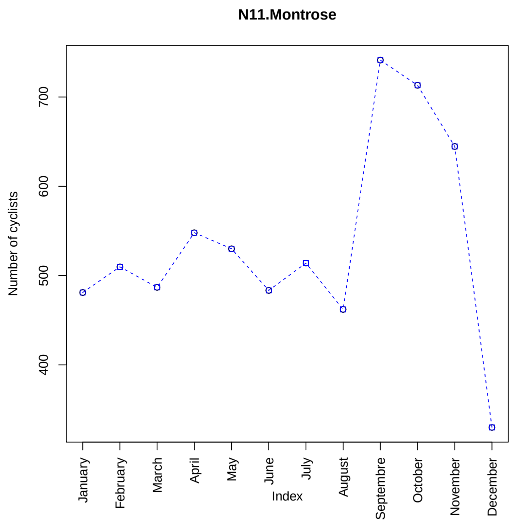

Figure 16: N11 Montrose two-way

Examining the in and out counts, the in count is less stable than the away count,

it has a greater fall in cyclist count in May, June, July and August than the out count. Again this can be explained by the demographic of the cyclists using each cycling lane. The 'in' counter is positioned in such a way it is more likely to pick up those people, students of UCD, that decide to use the cycle lane. Students turning left into UCD are more likely to use the bicycle lane (on the left of the road) than commuters going further than UCD who are more likely to avoid the cycle lane as taking the cycle lane would slow them down (and for reasons stated above). As a result when the proportion of students cycling on the N11 Montrose cycle lane towards town falls in May, June, July, August and December as they are in exam/holiday period, the total cyclist count falls in those same months.
The fall in cyclists for the out away from town counter, is not as large due to the fact that there are a greater proportion of town commuters making use of the cycle lane on that side of the N11.

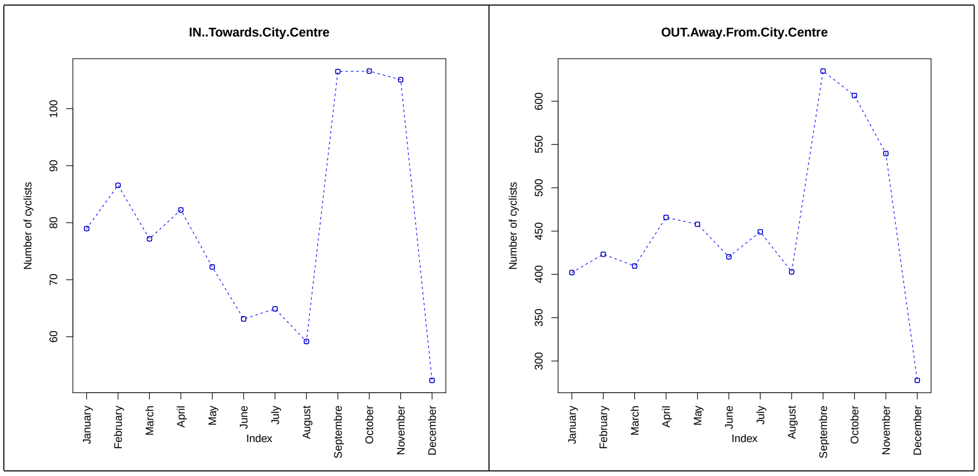

Figure 17: N11 Montrose In and Out

Clonskeagh (see Figure 18 and 19): A similar explanation to the one produces above can be given for the Clonskeagh road location. Quite a large proportion of students would commute through the Clonskeagh road, accounting for the fall in cyclists in June, July, August and December and the large upward jump in September. The trend differs in that we see a low in March followed by a marked increase in April. We will abstain from hypothesising on this data as there is a lot of missing dates.

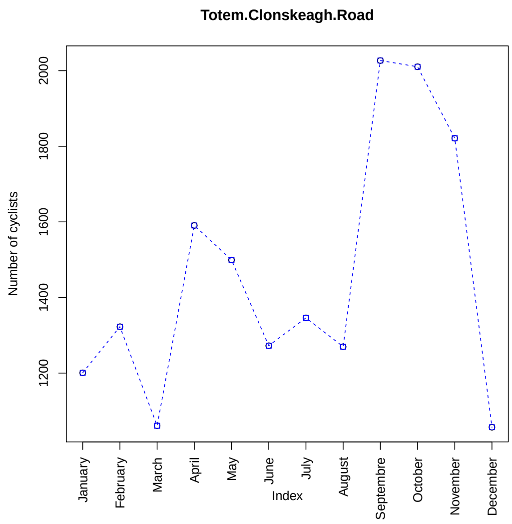

Figure 18: Clonskeagh Road

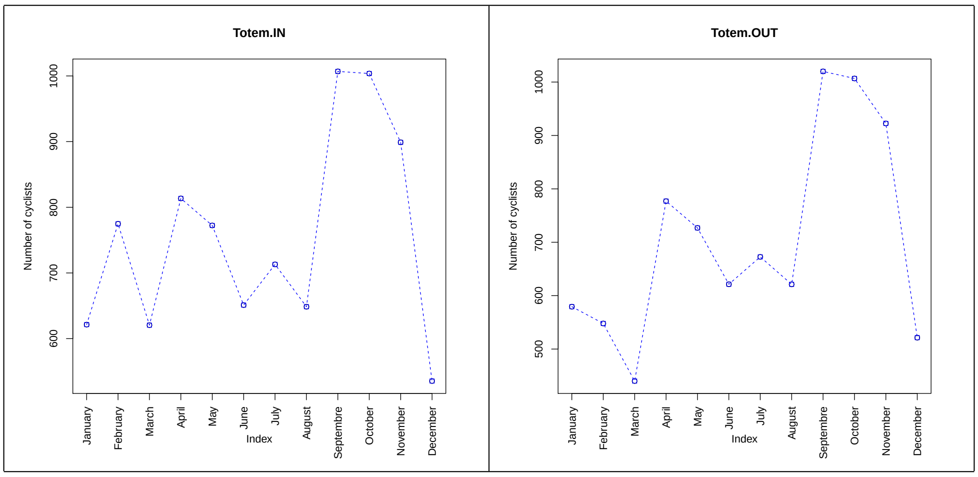

Figure 19: Clonskeagh Road In and Out

N11 Stillorgan (see Figure 20):

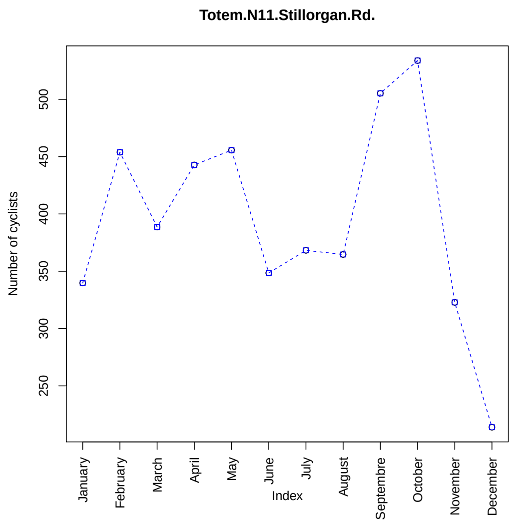

Figure 20: N11 Stillorgan Road

2012-2014 Mean Monthly Count

Listing 16: 2012-2014 Mean Monthly Count

```
trainingmonthyear <-apply(cycle_training[,2:16], 2,
   function(x) tapply(x, list(cycle_training$Month,
   cycle_training$Year), mean,na.rm=TRUE))
training2monthyear <-apply(cycle_training2[,2:21], 2,
   function(x) tapply(x, list(cycle_training2$Month,
   cycle_training2$Year), mean,na.rm=TRUE))

for (i in (1:15)){
  pdf(paste("trainingmeants3", i,".pdf",sep=""))
  plot(trainingmonthyear[,i],ylab='Number of cyclists',
     main=names(cycle_training[2:16])[i],xaxt='n', lty
     =3)
  axis(1, at=1:36, labels=c('','',"March '12",'','',"
     June '12",'','',"Septembre '12",'','',"December
     '12",'','',"March '13",'','',"June '13",'','',"
     Septembre '13",'','',"December '13",'','',"March
     '14",'','',"June '14",'','',"Septembre '14",'','',"
     December '14"),las=2)
  lines(trainingmonthyear[,i], type="o", pch=22, lty=2,
     col="red")
  dev.off()}
```

```
for (i in (1:20)){
  pdf(paste("training2meants3", i,".pdf",sep=""))
  plot(training2monthyear[,i],ylab='Number of cyclists',
     main=names(cycle_training2[2:21])[i],xaxt='n', lty
     =3)
  axis(1, at=1:36, labels=c('','',"March '12",'','',"
     June '12",'','',"Septembre '12",'','',"December
     '12",'','',"March '13",'','',"June '13",'','',"
     Septembre '13",'','',"December '13",'','',"March
     '14",'','',"June '14",'','',"Septembre '14",'','',"
     December '14"),las=2)
  lines(training2monthyear[,i], type="o", pch=22, lty=2,
      col="red")
  dev.off()}
```

The three year averaged monthly mean trends presented in the previous section demonstrated that cyclist count was clearly affected by seasonal effects, with different locations affected differently depending on the location of the counter and the demographic of the cyclists on that route. The monthly mean over the 2012-2014 time series presented in this section reveals an additional element of growth over the years. The peaks and throughs of the trends are higher year after year, indicating that the number of cyclists has been increasing over the 2012-2014 period.

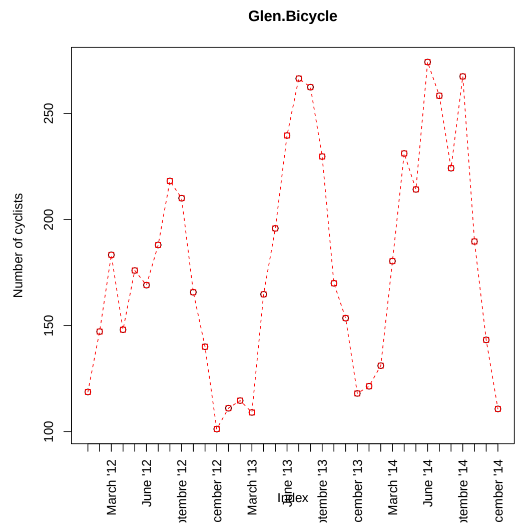

Figure 21: The Metals two-way (Bicycle only)

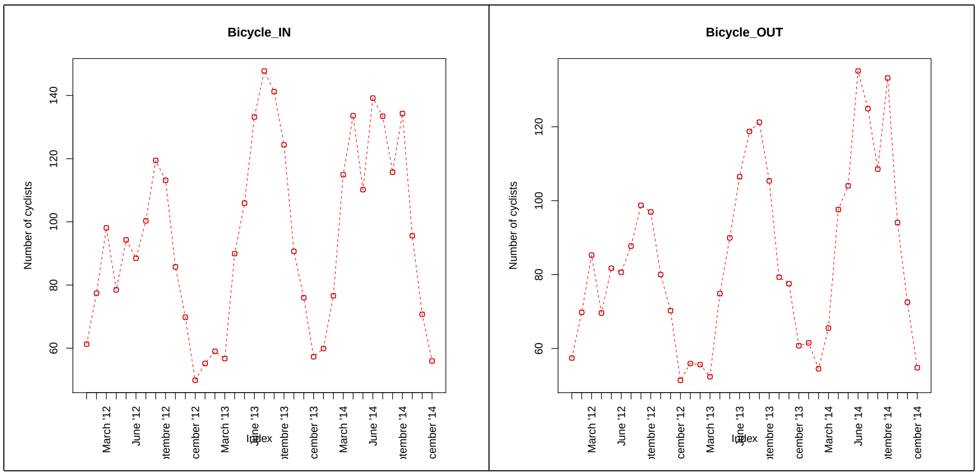

Figure 22: The Metals In and Out

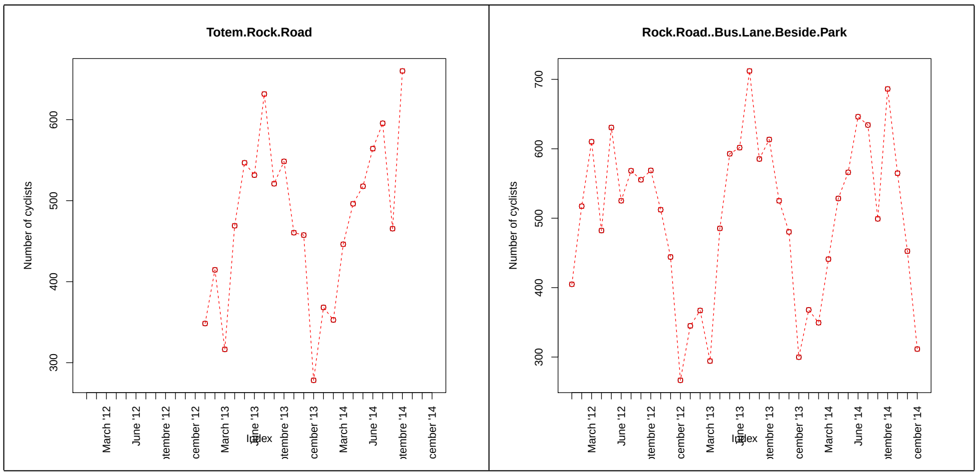

Figure 23: Totem Rock Road and Rock Road Bus Lane Beside Park

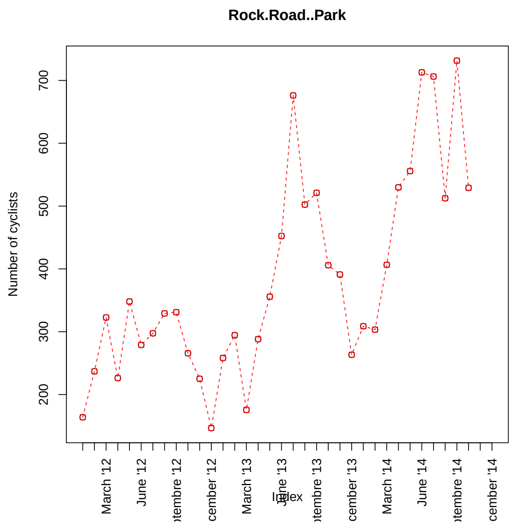

Figure 24: Rock Road Park two-way

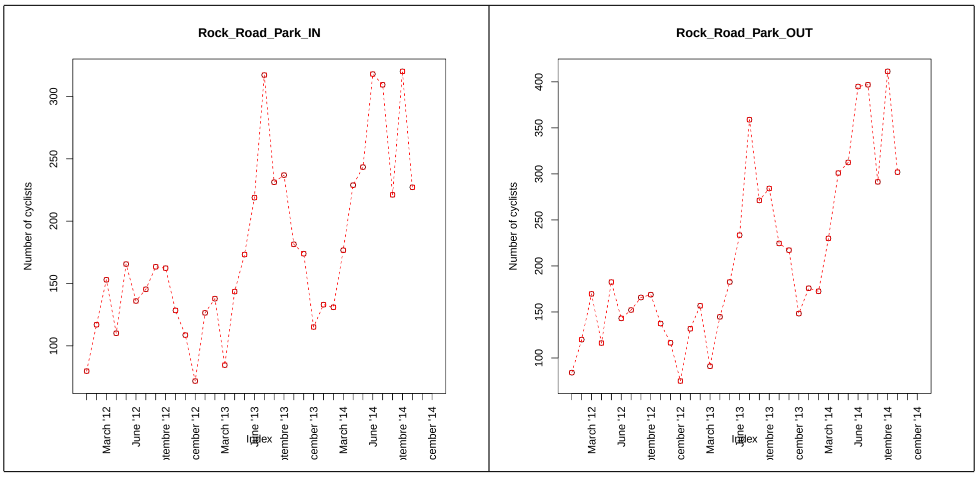

Figure 25: Rock Road Park In and Out

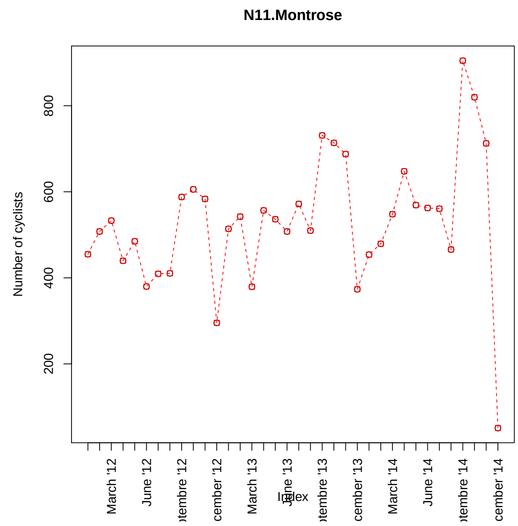

Figure 26: N11 Montrose two-way

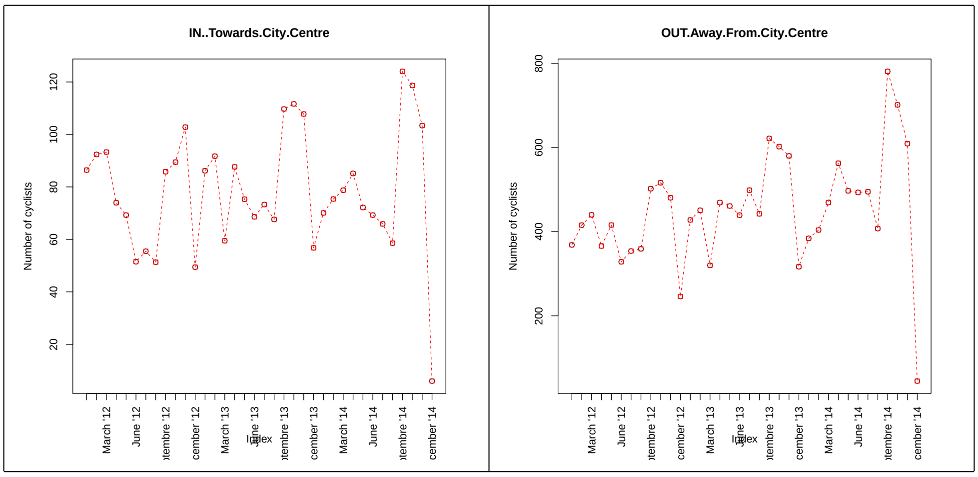

Figure 27: N11 Montrose In and Out

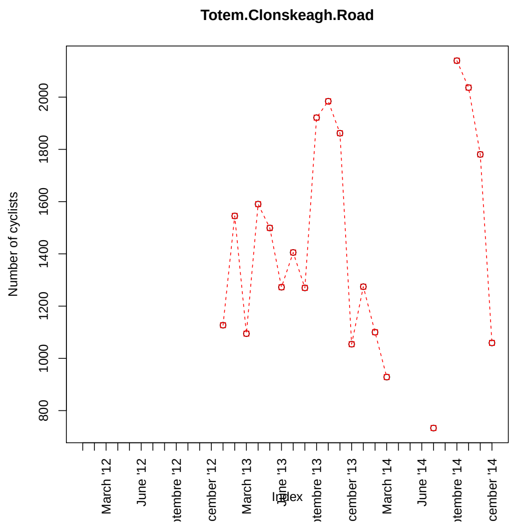

Figure 28: Clonskeagh Road

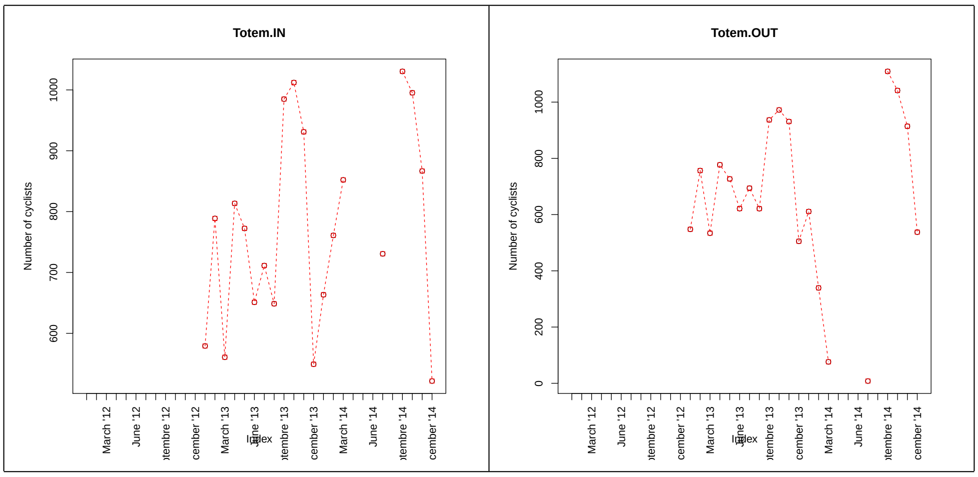

Figure 29: Clonskeagh Road In and Out

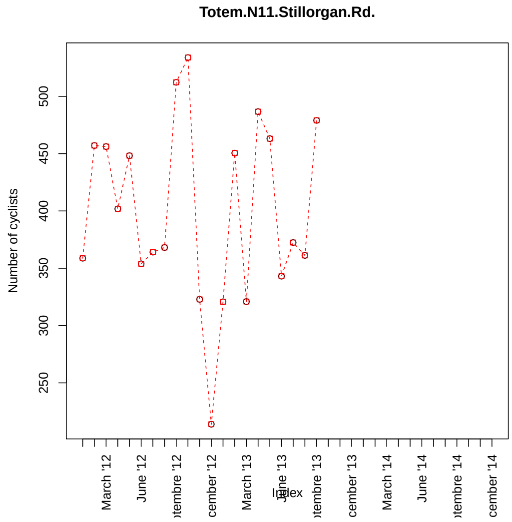

Figure 30: N11 Stillorgan Road

Conclusion

In the first section of this project, I pre-processed the bicycle counter data to create one single dataset for the three years by removing those locations which were not present in the yearly data set with the lowest number of locations (2012). Tabular and histogram summaries of the data were created to get a better perspective of the data's distribution. Next I plotted time series graphs of the mean monthly data. Analysing the time-series plots, I decided to revert to the data pre-processing step of the CRISP-DM framework as I wanted to explore some of the locations I had exempted form the first dataset. A new dataset was created which included all locations. Missing data was assigned with NA, which would be ignored in mean calculations. Next we went back to the 'data understanding' step of the process to conclude that we were dealing with seasonal data, affected in different ways depending on the location of the counter.

We could say that the locations' trend is split into two categories, one where the seasonal effect is distributed with a bell shaped (where the peak of the bell is during the summer months), and another 'left-skewed' distribution were bicycle count is stable for the first 8 months of the year and booms in September to fall to a trough in December.

2. Predict the number of cyclists each month at each counter site using the data from 2012, 2013 and 2014 as training data, and 2015 as a test set.

(a) Describe the steps undertaken to achieve this task.

Having pre-processed and understood the data, we can move on to modelling. Here we will try and achieve our objective, to predict the number of cyclists each at each counter site using 2012-2014 data as a training dataset and 2015 as a test set.

We are dealing with seasonal data which is trending upward over time (cycling count growing over time). A model that could predict this seasonal effect over time would need to separate out the upward trend over time and the seasonal effect.

The Holt-Winters seasonal method will be used to capture the seasonality of the data. The method comprises of a forecast equation and three smoothing equations (for the level, trend and seasonality). We provide a monthly seasonality component (m=12). There are two variations of Holt-Winters, the additive and multiplicative methods. The former is preferred when seasonal variations are roughly constant over the series while the multiplicative method is preferred where seasonal changes are proportional. Each variation may be better suited to different locations. For example, the multiplicative method may suit the 'Rock Road' locations (see Figure 24 and 25) better as the level of the seasonal effect increases over the 2012-2014 period, whereas the additive method may suit 'The Metals' better as there is less change in the level of the seasonality (see Figures 21, 22 and 23). For simplicity, we will pursue the additive model for all locations.
To begin our forecasting we revert to the data preparation step of the CRISP-DM framework. We split our mean monthly dataset ranging from January 2012 to December 2015 into training and test datasets (see listing 17). 'Totem N11 Stillorgan Road' and 'Totem Rock Road' will be excluded from the forecast as the test data is missing for both locations.
Then we move back to the modelling phase. We will use the monthly mean training dataset from January 2012 to December 2014 to train our Holt-Winters model and forecast the mean monthly count for the next 12 months (2015) using the trained model. Lastly, we will use the test dataset to measure the accuracy of our Holt-Winters forecast (see listing 18).

Listing 17: Splitting the monthly mean 2012-2015 dataset into two separate training and test datasets

```
train_x <- ts(test, frequency=12,start=c(test$Year[1],
   test$Month[1]))
test_x <- window(train_x, start=c(2015, 1))
train_x<- window(train_x, end=c(2014, 12))
```

Listing 18: Modelling

```
RRBus_HW<-HoltWinters(train_x[,9])
RRBus_F_HW <- forecast.HoltWinters(RRBus_HW, h=12)
accuracy(RRBus_F_HW, test_x[,9])
```

(b) Present an analysis of the results highlighting any limitations or weaknesses.

We train our Holt-Winters model, forecast for the next 12 months (2015) and lastly test the forecast's accuracy against the 2015 data (see example for N11 Montrose in listing 19).

Listing 19: N11 Montrose Training and Forecasting

```
pdf("HWN11.pdf")
plot(HoltWinters(train_x[,6]),main="Holt-Winters N11
   Montrose")
dev.off()
pdf("HWN11OUT.pdf")
plot(HoltWinters(train_x[,7]),main="Holt-Winters N11
   Montrose OUT")
dev.off()
pdf("HWN11IN.pdf")
plot(HoltWinters(train_x[,8]),main="Holt-Winters N11
   Montrose IN")
dev.off()

N11_HW<-HoltWinters(train_x[,6])
N11OUT_HW<-HoltWinters(train_x[,7])
N11IN_HW<-HoltWinters(train_x[,8])

N11_F_HW <- forecast.HoltWinters(N11_HW, h=12)
N11IN_F_HW <- forecast.HoltWinters(N11IN_HW, h=12)
```

```
N11OUT_F_HW <- forecast.HoltWinters(N11OUT_HW, h=12)

pdf("HWN11F.pdf")
plot(N11_F_HW,main="Holt-Winters Forecast for N11
   Montrose")
lines(test_x[,6], col='red')
dev.off()

pdf("HWN11INF.pdf")
plot(N11IN_F_HW,main="Holt-Winters Forecast for N11
   Montrose IN")
lines(test_x[,8], col='red')
dev.off()

pdf("HWN11OUTF.pdf")
plot(N11OUT_F_HW,main="Holt-Winters Forecast for N11
   Montrose OUT")
lines(test_x[,7], col='red')
dev.off()

xtable(accuracy(N11_F_HW, test_x[,6]))
xtable(accuracy(N11IN_F_HW, test_x[,8]))
xtable(accuracy(N11OUT_F_HW, test_x[,7]))
```

The Metals Glenageary

The plot of the observed data (black line) versus fitted points (red line) in Figures (31, 32 and 33) demonstrate that the Holt-Winters model for the Glenageary location fits the training data quite well. Table 17, 18 and 19 allow us to objectively measure the model's accuracy on the training set (first row). The Root Mean Squared Error (RMSE) allows to see the average error of the model which is only meaningful when compared to the magnitude of the counter values for that location. The Mean Absolute Scaled Error (MASE) is a better measure when comparing the accuracy across series. MASE indicates that our models for Glenageary are of similar accuracy.

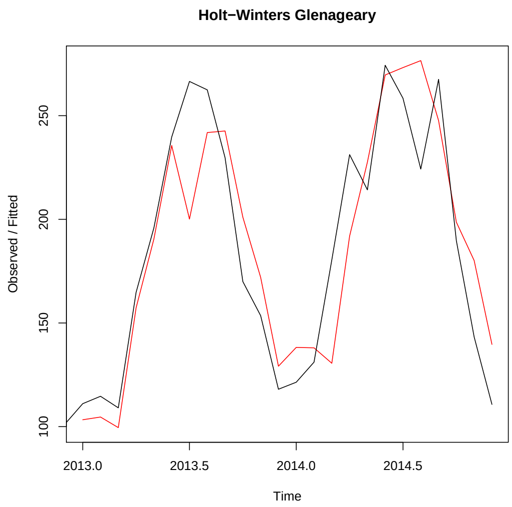

Figure 31: Holt-Winters Filtering Glenageary Two-Way

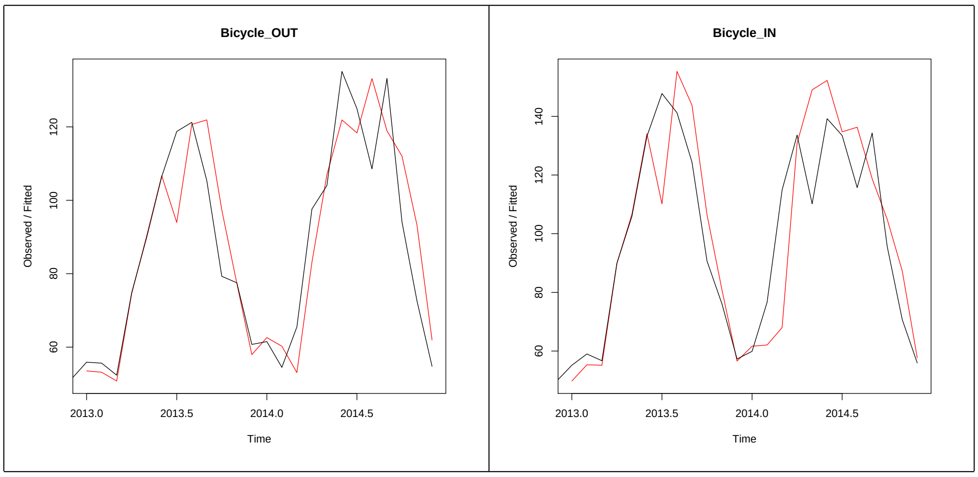

Figure 32: Holt-Winters Filtering Glenageary In and Out

The forecasted 2015 lines (blue line) for the three plots (Figures 33, 34, 35) seem quite accurate, bar the Glenageary Out forecast which is slightly underestimated.

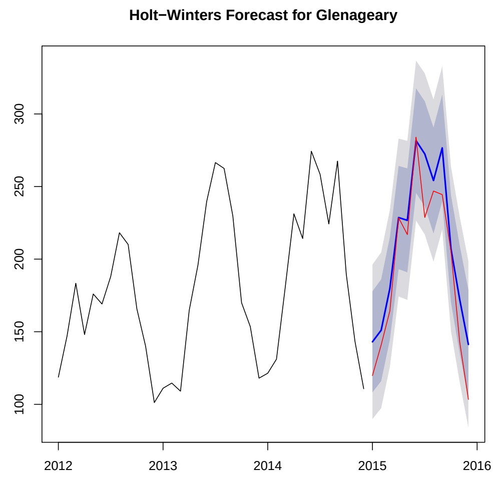

Figure 33: Holt-Winters Forecast Glenageary Two-Way

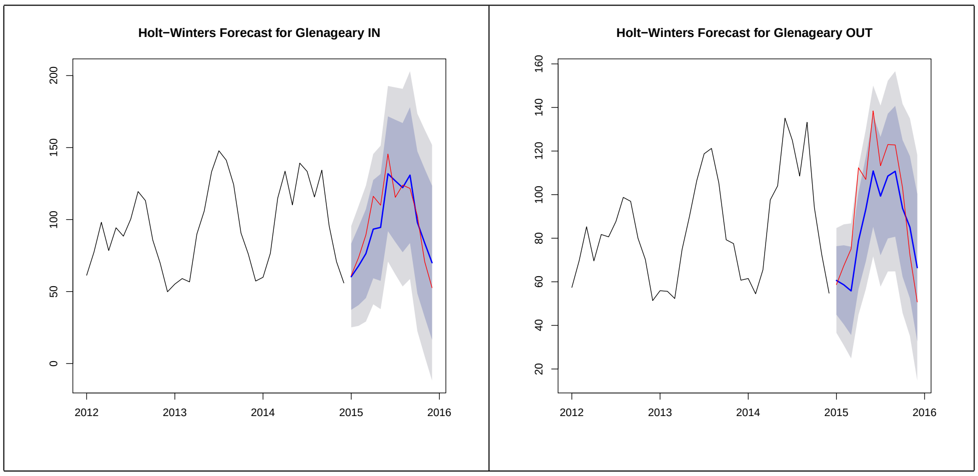

Figure 34: Holt-Winters Forecast Glenageary Two-Way In and Out

Examining the accuracy on the test data, Glenageary two-way is best (0.58) and Glenageary out does considerably worse than others (1.13) as we could expect examining the Figure 34.

Table 17: Glenageary two-way Accuracy

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-0.32	26.59	20.72	-0.10	12.93	0.62	0.30	
Test set	-17.36	22.87	17.75	-11.07	11.21	0.58	-0.31	0.53

Table 18: Glenageary IN Accuracy

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-1.25	17.61	11.98	0.88	13.88	0.62	0.14	
Test set	2.22	12.44	10.72	0.31	11.77	0.60	0.43	0.66

Table 19: Glenageary OUT Accuracy

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-0.81	12.03	8.81	-0.76	10.68	0.63	0.05	
Test set	10.19	17.25	15.26	7.69	16.34	1.13	0.45	0.89

For all three model's, the Theil's U statistics indicates that the forecast would do better than we would have done by guessing future values.

N11 Montrose

The N11 Montrose models fit the training data relatively well, both the trend and seasonal changes of the data are represented. MASE for all three models is slightly higher than that which we found Glenageary, in particular the IN model (0.81).

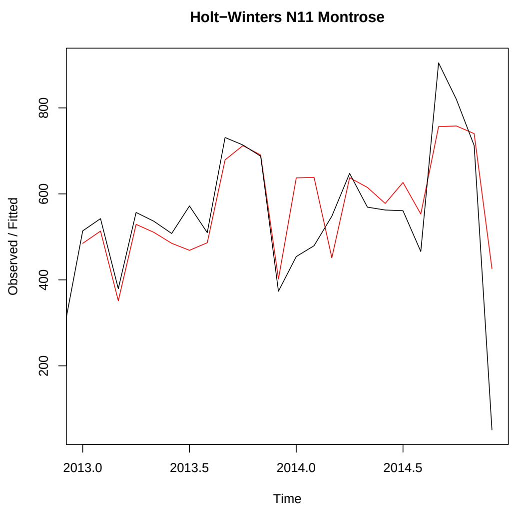

Figure 35: Holt-Winters Filtering N11 Montrose Two-Way

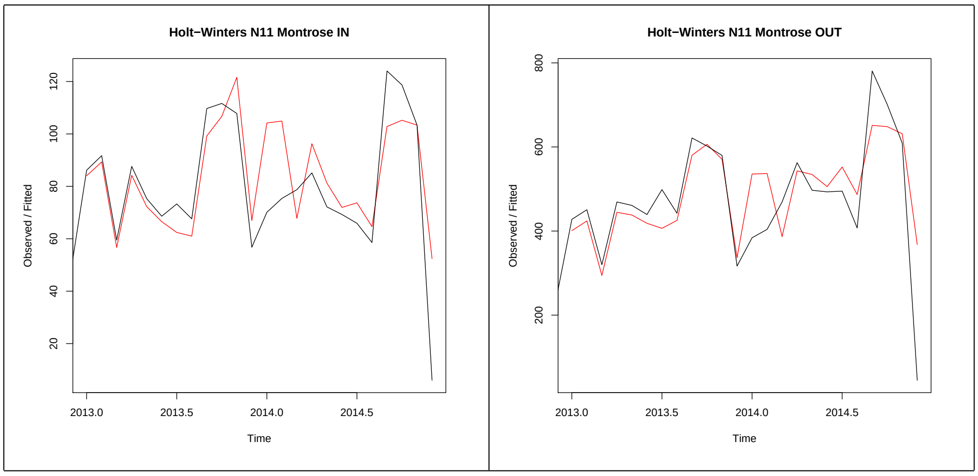

Figure 36: Holt-Winters Filtering N11 Montrose In and Out

The forecast for 2015 is also quite accurate, the blue line follows the test data's red line closely. Even though we are missing data for the first three months of 2015, we can see that in all three cases, the forecasted line meets the observed data line in April.

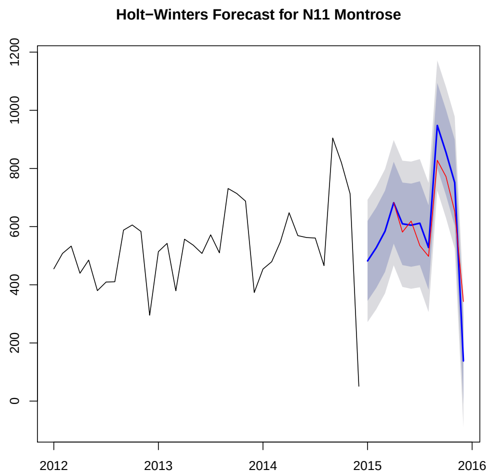

Figure 37: Holt-Winters Forecast N11 Montrose Two-Way

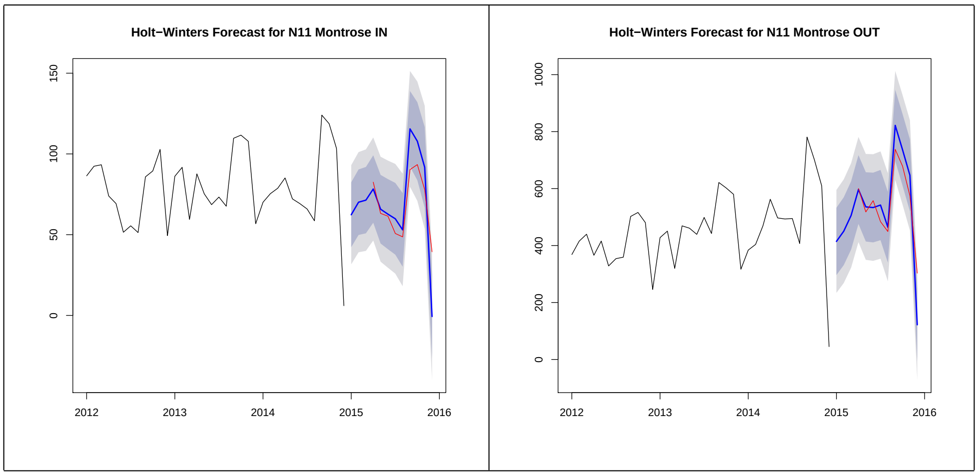

Figure 38: Holt-Winters Forecast N11 Montrose In and Out

Table 20: N11 Montrose two-way Accuracy

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-13.79	105.87	68.75	-3.69	14.62	0.67	0.16	
Test set	-24.29	94.70	72.93	-0.41	14.20		-0.11	

Table 21: N11 Montrose IN Accuracy

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-3.19	15.66	11.05	-4.79	14.88	0.81	0.26	
Test set	-2.94	17.55	12.79	1.42	22.32		-0.02	

Table 22: N11 Montrose OUT Accuracy

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-10.40	90.83	59.69	-3.40	14.95	0.67	0.15	
Test set	-10.60	76.70	57.03	1.51	12.88		-0.13	

The analysis has an important limitation, due to the fact that we have missing data for the first three months of 2015, we cannot calculate a MASE measure for the test set. A solution would be to shorten the test set and start the forecast from the April 2015.
Another weakness is the 'jump' we observe in all three cases from the end of observed line to the beginning of the forecast line.

Rock Road Bus Lane Beside Park

The model for 'Rock Road Bus Lane Beside Park' is considerably more accurate than previous models we have looked at (test MASE = 0.47).

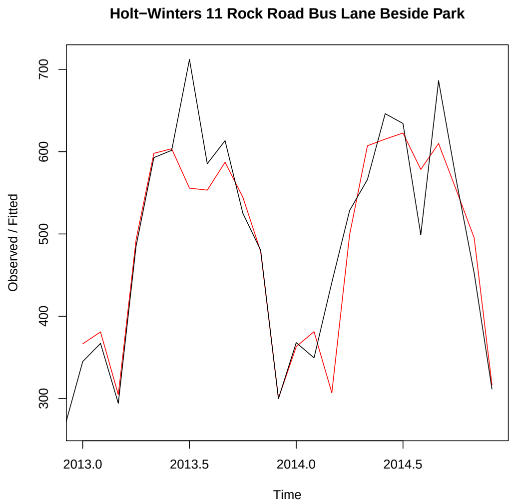

Figure 39: Holt-Winters Filtering Rock Road Bus Lane Beside Park

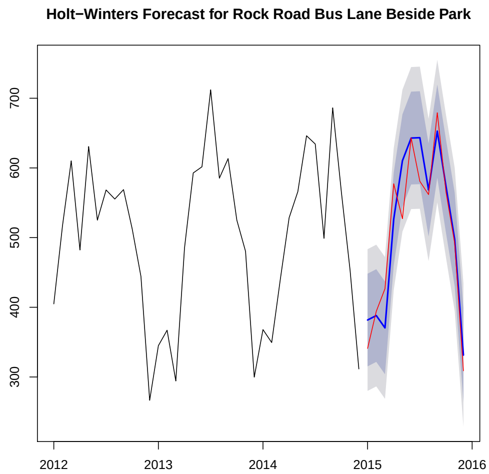

Figure 40: Holt-Winters Forecast Rock Road Bus Lane Beside Park

Table 23: Rock Road Bus Lane Beside Park Accuracy

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	9.76	51.72	33.17	2.35	7.04	0.42	-0.10	
Test set	-7.42	40.43	30.73	-1.83	6.41	0.47	-0.07	0.39

Rock Road Park two-way, In and Out

(a) When trying to run HoltWinters on the Rock Road Park location we get an error relating to NAs. There are two NAs for the last two months of the training dataset relating to the Rock Road Park location (November and December 2014). To avoid this problem, we change the training data's range from January 2012 to October 2014 and the test data set back from November 2014 to December 2015 (see listing 19). Thus, the training data set was shortened by two months, and the test dataset lengthened by the same amount.

Listing 20: Altering the training and test datasets to avoid missing data errors

```
train_x_park <- ts(test, frequency=12,start=c(test$Year
   [1], test$Month[1]))
test_x_park <- window(train_x_park, start=c(2014, 11))
train_x_park<- window(train_x_park, end=c(2014, 10))
```

(b)

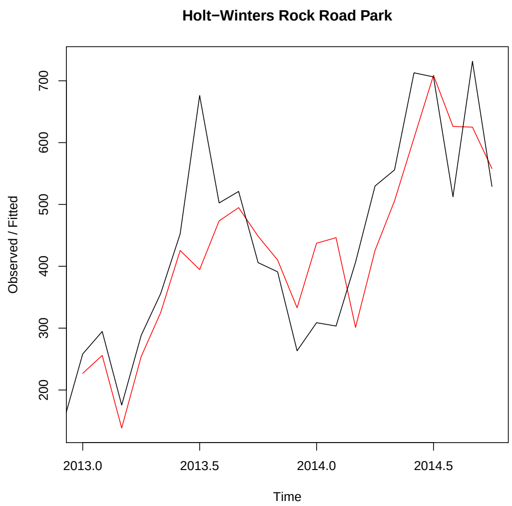

Figure 41: Holt-Winters Filtering Rock Road Park Two-Way

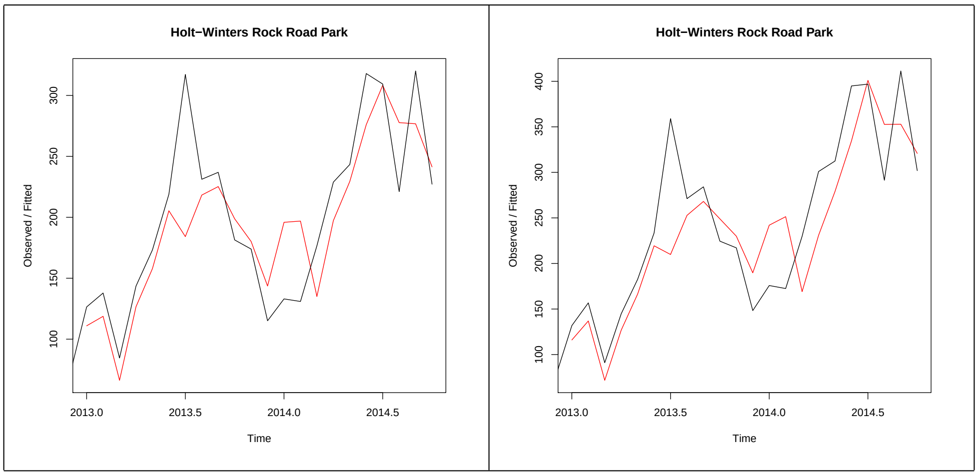

Figure 42: Holt-Winters Filtering Rock Road Park In and Out

(a) As we have shortened the training dataset by two months we must extend the forecast by two months (h=14 instead of h=12).

Listing 21: Forecasting 14 months

```
RRP_HW<-HoltWinters(train_x_park[,10])
RRPIN_HW<-HoltWinters(train_x_park[,11])
RRPOUT_HW<-HoltWinters(train_x_park[,12])

RRP_F_HW <- forecast.HoltWinters(RRP_HW, h=14)
RRPIN_F_HW <- forecast.HoltWinters(RRP_HW, h=14)
RRPOUT_F_HW <- forecast.HoltWinters(RRP_HW, h=14)
```

(b)
The test data is missing data in the first 5 months (November 2014 to March 2015), thus it is hard to say how accurate our model is. In all three cases, the model does not match the observed data for 2015, there is quite a large error. The model's weakness may be the result of having had to shorten the training data set.

Examining Figures 43 and 44, there is a much steeper bell shaped curved in 2013 than in 2012. When training on the training set, the model forecasts this upward long-term trend which results in bicycle counts being over-estimated in 2015.

A possible solution would be to exempt the 2012 training data. This would mean that the model would not need to fit such a steep upward 'growth' trend.
Alternatively, we could set a lower trend (b_t) parameter in the Holt-Winters model.

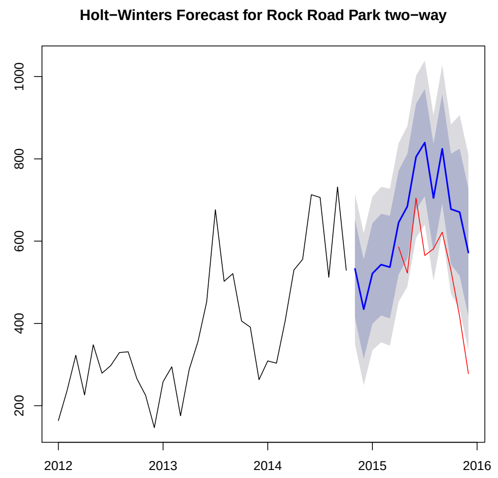

Figure 43: Holt-Winters Forecast Rock Road Park Two-Way

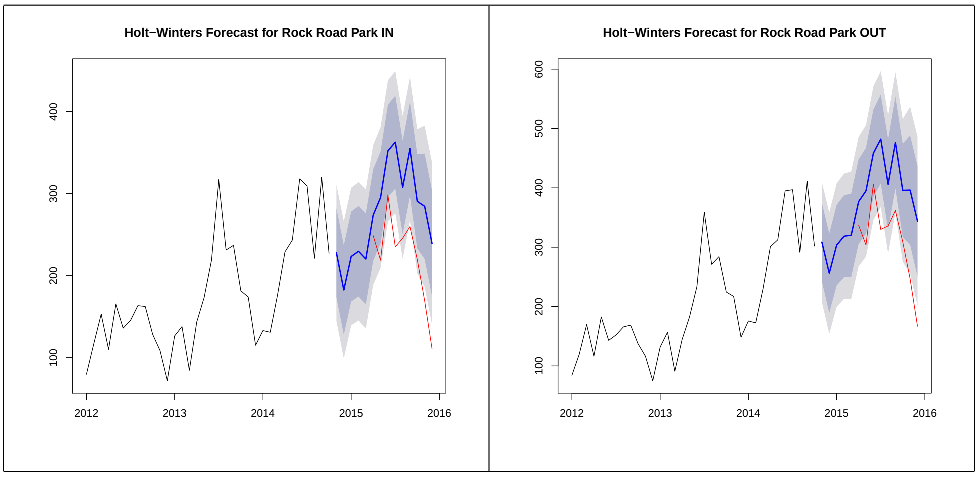

Figure 44: Holt-Winters Forecast Rock Road Park In and Out

Table 24: Rock Road Park two-way Accuracy

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	20.86	93.42	70.68	6.61	24.82	0.50	0.17	
Test set	-180.02	195.80	180.02	-39.22	39.22		-0.09	

Table 25: Rock Road Park IN Accuracy

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	8.10	42.14	30.98	5.33	22.49	0.49	0.16	
Test set	-83.82	90.22	83.82	-43.88	43.88		-0.10	

Table 26: Rock Road Park Out Accuracy

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	11.88	51.61	39.90	7.38	27.10	0.50	0.20	
Test set	-103.70	113.06	103.70	-38.67	38.67		-0.04	

Clonskeagh two-way, In and Out

(a) Training data for the 'Clonskeagh Road' location is missing for the whole of 2012 and for some of 2014. The Holt-Winters model requires at least two period (two years in our case) to train on. We copy the 2013 data onto the 2012 missing data to extend our dataset (see listing 22).

Listing 22: Copying 2013 data into 2012 missing values

```
train_x_clon[1,16] <-train_x_clon[13,16]
train_x_clon[2,16] <-train_x_clon[14,16]
train_x_clon[3,16] <-train_x_clon[15,16]
train_x_clon[4,16] <-train_x_clon[16,16]
train_x_clon[5,16] <-train_x_clon[17,16]
train_x_clon[6,16] <-train_x_clon[18,16]
train_x_clon[7,16] <-train_x_clon[19,16]
train_x_clon[8,16] <-train_x_clon[20,16]
train_x_clon[9,16] <-train_x_clon[21,16]
train_x_clon[10,16] <-train_x_clon[22,16]
train_x_clon[11,16] <-train_x_clon[23,16]
train_x_clon[12,16] <-train_x_clon[24,16]

train_x_clon[1,17] <-train_x_clon[13,17]
train_x_clon[2,17] <-train_x_clon[14,17]
train_x_clon[3,17] <-train_x_clon[15,17]
train_x_clon[4,17] <-train_x_clon[16,17]
train_x_clon[5,17] <-train_x_clon[17,17]
train_x_clon[6,17] <-train_x_clon[18,17]
train_x_clon[7,17] <-train_x_clon[19,17]
```

```
train_x_clon[8,17] <-train_x_clon[20,17]
train_x_clon[9,17] <-train_x_clon[21,17]
train_x_clon[10,17] <-train_x_clon[22,17]
train_x_clon[11,17] <-train_x_clon[23,17]
train_x_clon[12,17] <-train_x_clon[24,17]

train_x_clon[1,18] <-train_x_clon[13,18]
train_x_clon[2,18] <-train_x_clon[14,18]
train_x_clon[3,18] <-train_x_clon[15,18]
train_x_clon[4,18] <-train_x_clon[16,18]
train_x_clon[5,18] <-train_x_clon[17,18]
train_x_clon[6,18] <-train_x_clon[18,18]
train_x_clon[7,18]<-train_x_clon[19,18]
train_x_clon[8,18] <-train_x_clon[20,18]
train_x_clon[9,18] <-train_x_clon[21,18]
train_x_clon[10,18] <-train_x_clon[22,18]
train_x_clon[11,18] <-train_x_clon[23,18]
train_x_clon[12,18] <-train_x_clon[24,18]
```

The 2014 missing data proves problematic again. We use the same method as in the previous section and ignore it by including the data beyond March 2014 as test data (see listing 23).

(b)

As expected, the fitted line exactly follows the observed line as we copied the 2013 data instead of the 2012 data. We can therefore expect our forecasted line to have the same trend as the line for 2012.

Listing 23: Setting March 2014 to December 2015 as test

```
train_x_clon <- ts(test, frequency=12,start=c(test$Year
   [12], test$Month[1]))
test_x_clon <- window(train_x_clon, start=c(2014, 4))
train_x_clon<- window(train_x_clon, end=c(2014, 3))
```

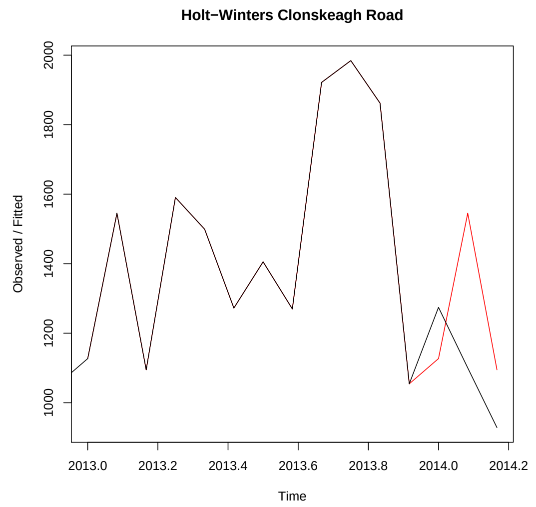

Figure 45: Holt-Winters Filtering Clonskeagh Two-Way

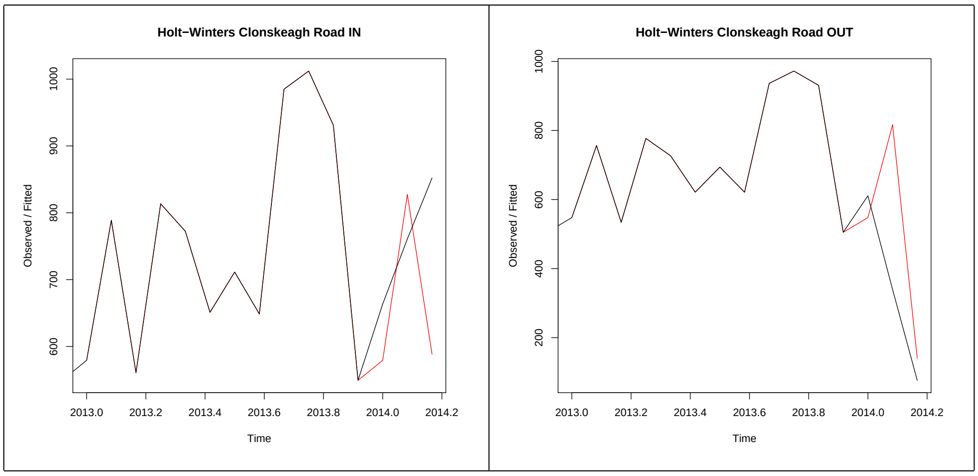

Figure 46: Holt-Winters Filtering Clonskeagh Road In and Out

(a) As we shortened the training dataset and lengthened the test dataset, we must increase the forecasting period from 12 months (h=12) to 21 months (h=21) (see listing 24).

(a) The model does not do a bad job at forecasting the two-way test data, the seasonal effect is similar in both lines. Clonskeagh in and out are not as successful in their forecast.
The trend in the Clonskeagh in forecast would have to be removed when modelling by decreasing the trend parameter.

In the out case, the seasonal trend seems accurate but the fall in cyclist count in the beginning of 2014 results in the forecasted trend being too low. The model level parameter would have to be adjusted.

Listing 24: Modelling and Forecasting

```
clon_HW<-HoltWinters(train_x_clon[,16])
clonIN_HW<-HoltWinters(train_x_clon[,17])
clonOUT_HW<-HoltWinters(train_x_clon[,18])

clon_F_HW <- forecast.HoltWinters(clon_HW, h=21)
clonIN_F_HW <- forecast.HoltWinters(clonIN_HW, h=21)
clonOUT_F_HW <- forecast.HoltWinters(clonOUT_HW, h=21)
```

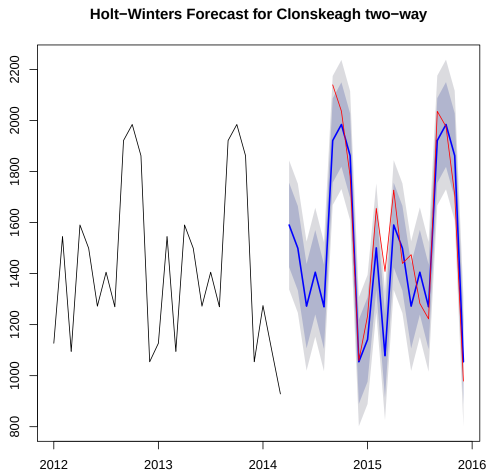

Figure 47: Holt-Winters Forecast Clonskeagh Two-Way

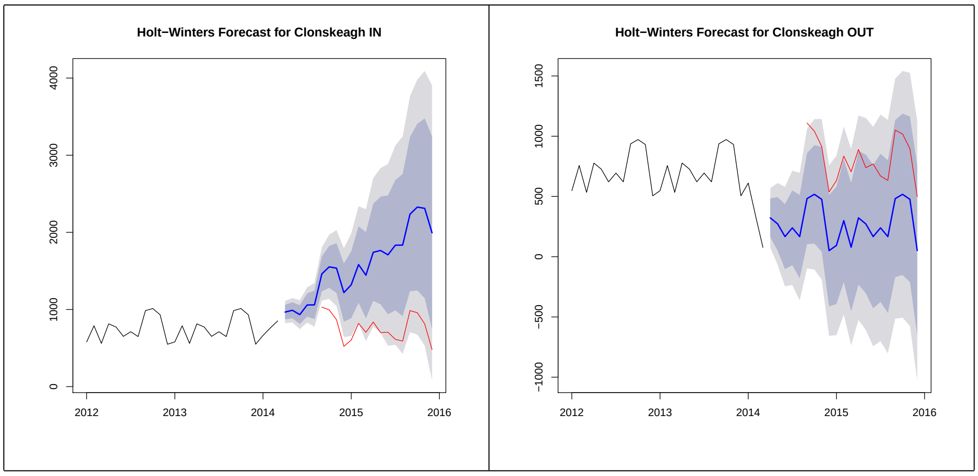

Figure 48: Holt-Winters Forecast Clonskeagh Road In and Out

Table 27: Clonskeagh Road two-way Accuracy

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-30.93	128.45	50.59	-2.06	3.81	Inf	-0.01	
Test set	46.99	148.34	124.67	2.74	8.51	2.46	0.21	0.35

Table 28: Clonskeagh Road IN Accuracy

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	18.78	73.56	27.65	3.55	4.67	Inf	-0.31	
Test set	-1107.34	1140.71	1107.34	-158.26	158.26	41.17	0.75	5.58

Table 29: Clonskeagh Road OUT Accuracy

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-31.85	125.40	40.27	-4.23	5.77	Inf	-0.06	
Test set	514.24	518.45	514.24	68.45	68.45	8.22	0.05	2.39

(c) Would you use your best predictive model? Discuss your answer.

The second last step of the CRISP DM framework is to evaluate our model before deciding whether to deploy it or to go back to the previous phases.

The main weakness of the models produced in this report are the result of missing data. We used two methods in dealing with missing data; reducing the size of the training data (N11 Montrose and Rock Road Park locations) or, if a whole year is missing by copying a year's values instead of another year's missing values (does not take trends into account). There are other methods for filling missing data which we have not pursued such as extrapolation or assigning a missing month's values with the mean of the two month's values closest to it.
By reducing the size of the training data, we reduce the 'experience' of the model. This can affect the model's ability to reproduce the seasonal effect of the data and the level of the data. Copying data from one year to the previous year, reduces the model's ability to take account for trends, luckily this did not prove problematic in the case of Clonskeagh road.

This report's best predictive model is the one created for the 'Rock Road Bus Lane Beside Park' location. The forecast has a low MASE of 0.47. Both the Glenageary models and the 'Rock Road Bus Lane Beside Park' models are not plagued by missing data. However, the latter preforms better than the former. This could be due to the fact that the trend for all three training years have a very similar shape, level and lack of trend. These factors increase the likelihood of the model forecasting well.

I would use the Holt-Winter's model for 'Rock Road Bus Lane Beside Park' as it is quite accurate and has a low Theil's U (0.39) measure meaning that it has good predictive capabilities. However I would be weary of using the Holt-Winters model on other locations for which the accuracy were not as good. The shape of the training data's time-series, its trend, whether there are missing values (and if so to what extent) and the model's ability to fit the training data would be factors I would take into consideration when deciding whether to use Holt-Winters to predict bicycle counter data into the future.
The model is successful in cases where the trend and seasonal effect are constant (for example Glenageary, N11 Montrose, Rock Road Beside Park). In such locations one can be confident in using Holt-Winters with the right parameters.
In the case of the Clonskeagh location I would not have used the model as there is too much missing data, there were original observable values for just over 12 months. However the model ended up being quite accurate, probably due to the fact that there was not much growth in the trend.
If the data's trend is inconsistent, as is the case with the Rock Road Park location (large growth form 2012 to 2013, but trend levels out there after), I would avoid using the model as it is likely to ovepredict or underpredict.