# Snowy winters

## a historical study of Toronto and area

Bill Gough Ken Butler

February 26, 2014

Section 1

# Introduction

# Introduction

- Data were collected by Bill Gough from 1850–2010.
- Each year, snow-to-rain ratio calculated for December and January; results averaged by decade.
- In addition, each winter was categorized as “snowy”, “rainy” or “mixed”; these also summarized by decade.

# Data

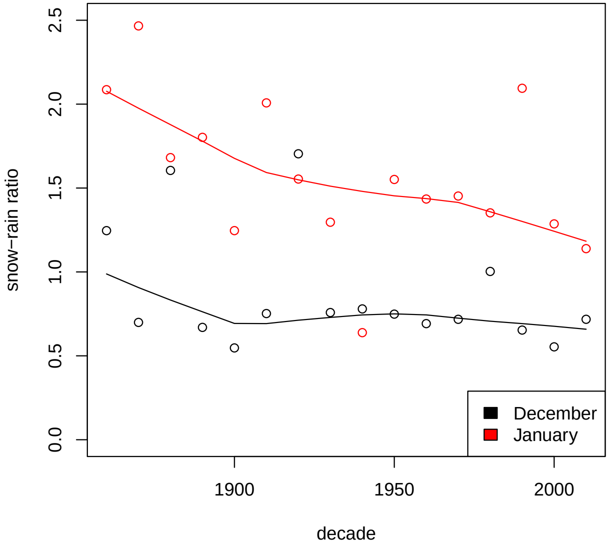
```
> snow=read.table("snow.txt",header=T)
> snow
```

| | decade | dec.snow.rain | jan.snow.rain | rainy | snowy | mixed |
|---|---|---|---|---|---|---|
| 1 | 1860 | 1.2461911 | 2.086197 | 1 | 3 | 6 |
| 2 | 1870 | 0.6991449 | 2.466387 | 1 | 3 | 6 |
| 3 | 1880 | 1.6049055 | 1.681356 | 2 | 4 | 4 |
| 4 | 1890 | 0.6695511 | 1.802118 | 2 | 4 | 4 |
| 5 | 1900 | 0.5472127 | 1.246255 | 3 | 1 | 6 |
| 6 | 1910 | 0.7520022 | 2.007601 | 2 | 1 | 7 |
| 7 | 1920 | 1.7041667 | 1.553692 | 2 | 5 | 3 |
| 8 | 1930 | 0.7580128 | 1.296309 | 2 | 3 | 5 |
| 9 | 1940 | 0.7796400 | 0.638236 | 4 | 3 | 3 |
| 10 | 1950 | 0.7485876 | 1.551166 | 3 | 2 | 5 |
| 11 | 1960 | 0.6919866 | 1.434611 | 2 | 2 | 6 |
| 12 | 1970 | 0.7176566 | 1.452458 | 2 | 2 | 6 |
| 13 | 1980 | 1.0029203 | 1.352194 | 3 | 3 | 4 |
| 14 | 1990 | 0.6535492 | 2.094637 | 3 | 2 | 5 |
| 15 | 2000 | 0.5535900 | 1.286375 | 1 | 0 | 9 |
| 16 | 2010 | 0.7177549 | 1.138889 | 3 | 1 | 6 |

Section 2

# Snow-rain ratios

# Plot of evolution over time



# The code for the plot

```
> attach(snow)
> plot(decade,dec.snow.rain,ylim=c(0,2.5),
+      ylab="snow-rain ratio")
> lines(lowess(decade,dec.snow.rain))
> points(decade,jan.snow.rain,col="red")
> lines(lowess(decade,jan.snow.rain),col="red")
> legend("bottomright",legend=c("December","January"),
+        fill=c("black","red"))
```

Lowess curve is smooth curve through trend (not affected by outliers).

# Testing for trend

- January ratios almost always higher than December ones.
- January trend appears decreasing.
- December trend appears steady.
- Statistically significant? Mann-Kendall test for trend, based on Kendall correlation with time (not affected much by vertical outliers).

```
> library(Kendall)
> MannKendall(dec.snow.rain)
```

```
tau = -0.217, 2-sided pvalue =0.26035
```

```
> MannKendall(jan.snow.rain)
```

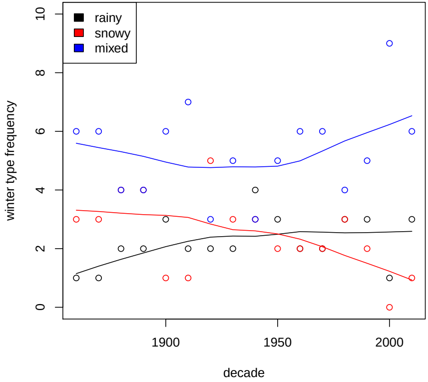
```
tau = -0.417, 2-sided pvalue =0.027377
```

- The January trend significant, but December one clearly not.

Section 3

# Rainy, snowy and mixed winters

- Each winter classified as being rainy, snowy or mixed, and the results summarized by decade.
- The number of years of each type in each decade must sum to 10
- but at first act as if independent.




- Definite downward trend in number of snowy winters per decade
- Trends of other types of winter less clear.

# Code for the plot

```
> plot(decade,rainy,ylim=c(0,10),
+      ylab="winter type frequency")
> lines(lowess(decade,rainy))
> points(decade,snowy,col="red")
> lines(lowess(decade,snowy),col="red")
> points(decade,mixed,col="blue")
> lines(lowess(decade,mixed),col="blue")
> legend("topleft",legend=c("rainy","snowy","mixed"),
+        fill=c("black","red","blue"))
```

# Mann-Kendall again

- Significance of trends confirmed by Mann-Kendall tests for each:

```
> MannKendall(rainy)
tau = 0.354, 2-sided pvalue =0.090938
> MannKendall(snowy)
tau = -0.402, 2-sided pvalue =0.045753
> MannKendall(mixed)
tau = 0.111, 2-sided pvalue =0.60673
```

- "Snowy" trend significantly downward (just, at 0.05 level of significance). "Rainy" trend significant at 0.10 level, but "mixed" trend not significant at all.
- Strictly speaking, should adjust for the fact that we've just done 3 tests.

Section 4

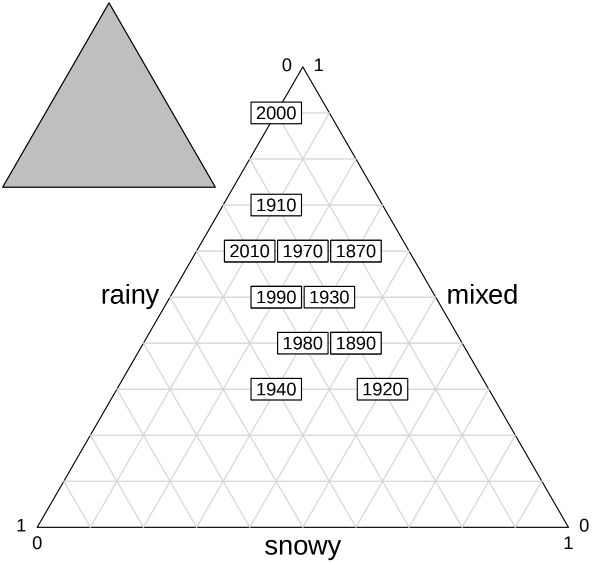
# Triangle plot of winter types

# What is a triangle plot?

- Used to display three variables that add up to a fixed total.
- In our case, `rainy`, `snowy` and `mixed` must add up to 10.
- A point in the triangle has three coordinates that add up to 1 (10, rescaled).
- Function `triangle.plot` in the package `ade4` draws these, except that first we need a little organization:

```
> winter.type=data.frame(rainy,snowy,mixed)
> decade.char=as.character(decade)
> library(ade4)
> triangle.plot(winter.type,label=decade.char,clabel=1,
+               scale=F,min3=c(0,0,0),max3=c(1,1,1))
```

# The triangle plot



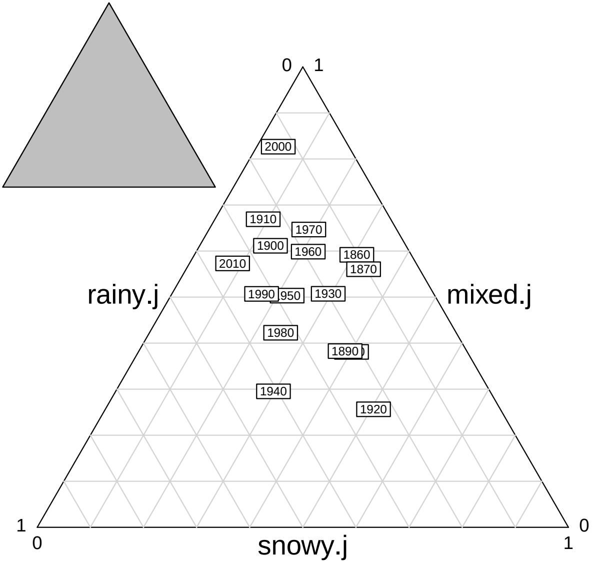
# Comments

- Not all decades shown because they overprint earlier decades. For example, 1880 and 1890 have identical profiles, and so do 1960 and 1970.
- As an example of how the plot works, 2000 had 9 mixed years, 1 rainy year and no snowy years, so it is plotted at 0.9 on the mixed scale, 0.1 on the rainy scale and 0.0 on the snowy scale. (The coordinates are a bit tricky to read off, but it can be done.)
- Likewise, 1920 is plotted at 0.2 on `rainy`, 0.5 on `snowy` and 0.3 on `mixed`.
- Wikipedia calls this a "ternary plot": see `http://en.wikipedia.org/wiki/Ternary_plot`.
- Trend over time would show up as eg. the early years occupying one part of the triangle and the late years another part. (I don't know a test for this.)
- Also, it would be nice to connect adjacent decades by lines, but I don't know how to make this happen on this triangle plot.

- Add random "jitter" to each observation, to make the hidden ones show up:

```
> rainy.j=jitter(rainy,amount=0.5)
> snowy.j=jitter(snowy,amount=0.5)
> mixed.j=jitter(mixed,amount=0.5)
> winter.j=data.frame(rainy.j,snowy.j,mixed.j)
> triangle.plot(winter.j,label=decade.char,clabel=1,
+                scale=F,min3=c(0,0,0),max3=c(1,1,1))
```

# Improved triangle plot



Section 5

# Data by individual year

# By year: organizing

```
> winter=read.csv("Winter Data _ Ken.csv",header=T)
```

```
> head(winter)
```

| | year | snow | rain | winter.type |
|---|---|---|---|---|
| 1 | 1873 | 223.1 | 38.2 | snowy |
| 2 | 1871 | 210.2 | 84.6 | mixed |
| 3 | 1867 | 180.5 | 105.3 | mixed |
| 4 | 1945 | 168.8 | 37.4 | snowy |
| 5 | 1860 | 165.5 | 79.5 | snowy |
| 6 | 1869 | 165.4 | 27.5 | snowy |

Arrange by year thus:

```
> o=order(winter$year)
> winter2=winter[o,]
> head(winter2)
```

| | year | snow | rain | winter.type |
|---|---|---|---|---|
| 96 | 1849 | 94.2 | 106.2 | rainy |
| 92 | 1850 | 97.4 | 85.4 | mixed |
| 87 | 1851 | 101.0 | 103.0 | mixed |
| 14 | 1852 | 154.3 | 44.5 | mixed |
| 85 | 1853 | 102.3 | 135.5 | mixed |
| 51 | 1854 | 121.4 | 86.0 | mixed |

# Rate evolution graph

- Cumulative occurrences of each winter type category against time. Weiss, http://www2.hsu-hh.de/mathstat/downloads/Folien_09_05.pdf.
- Separate plot on the same graph for each winter type.
- First, define logical variable: true (value 1) if winter of appropriate type, false (value 0) otherwise.

# Is it a rainy winter?

For rainy winters:

```
> attach(winter2)
> is.rainy=(winter.type=="rainy")
> head(data.frame(year,winter.type,is.rainy), n=12)
```

```
   year winter.type is.rainy
1  1849       rainy     TRUE
2  1850       mixed    FALSE
3  1851       mixed    FALSE
4  1852       mixed    FALSE
5  1853       mixed    FALSE
6  1854       mixed    FALSE
7  1855       snowy    FALSE
8  1856       snowy    FALSE
9  1857       mixed    FALSE
10 1858       rainy     TRUE
11 1859       mixed    FALSE
12 1860       snowy    FALSE
```

# Rainy winters so far

- To count the number of rainy winters so far, we use `cumsum` to cumulate the values in `is.rainy`, noting that `TRUE` has a numerical value of 1 and `FALSE` a value of 0:

```
> rainy.sofar=cumsum(is.rainy)
> head(data.frame(year,winter.type,is.rainy,rainy.sofar),n=12)
```

```
   year winter.type is.rainy rainy.sofar
1  1849       rainy     TRUE           1
2  1850       mixed    FALSE           1
3  1851       mixed    FALSE           1
4  1852       mixed    FALSE           1
5  1853       mixed    FALSE           1
6  1854       mixed    FALSE           1
7  1855       snowy    FALSE           1
8  1856       snowy    FALSE           1
9  1857       mixed    FALSE           1
10 1858       rainy     TRUE           2
11 1859       mixed    FALSE           2
12 1860       snowy    FALSE           2
```

# Same with mixed and snowy

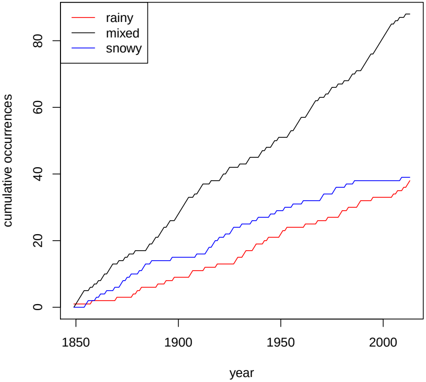
```
> is.mixed=winter.type=="mixed"
> mixed.sofar=cumsum(is.mixed)
> is.snowy=winter.type=="snowy"
> snowy.sofar=cumsum(is.snowy)
```

# Checking our work

```
> head(data.frame(year,winter.type,mixed.sofar,snowy.sofar,
+                  rainy.sofar),n=12)

   year winter.type mixed.sofar snowy.sofar rainy.sofar
1  1849       rainy           0           0           1
2  1850       mixed           1           0           1
3  1851       mixed           2           0           1
4  1852       mixed           3           0           1
5  1853       mixed           4           0           1
6  1854       mixed           5           0           1
7  1855       snowy           5           1           1
8  1856       snowy           5           2           1
9  1857       mixed           6           2           1
10 1858       rainy           6           2           2
11 1859       mixed           7           2           2
12 1860       snowy           7           3           2
```

# Our rate evolution graph



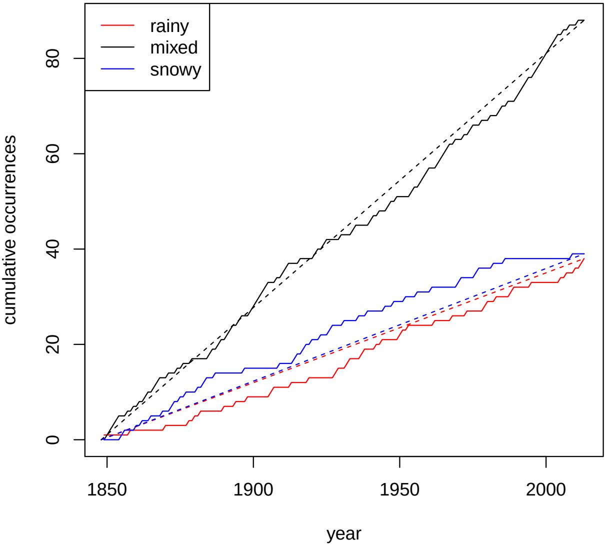

If proportion of winters of each type constant over time, traces should be straight.
"Snowy" seems to have stopped going up.

# Code for plot

```
> plot(mixed.sofar~year,type="n",
+       ylab="cumulative occurrences")
> points(rainy.sofar~year,col="red")
> points(mixed.sofar~year,col="black")
> points(snowy.sofar~year,col="blue")
> legend("topleft",c("rainy","mixed","snowy"),
+         col=c("red","black","blue"),lty="solid")
```

To guide eye, add suitably coloured line joining first and last points on trace.

# Rate evolution graph with lines



# The extra code

```
> year1=min(year)-1 #1848, no winters of each type
> year2=max(year)
> lines(c(year1,year2),c(0,max(rainy.sofar)),col="red",
+        lty="dashed")
> lines(c(year1,year2),c(0,max(mixed.sofar)),col="black",
+        lty="dashed")
> lines(c(year1,year2),c(0,max(snowy.sofar)),col="blue",
+        lty="dashed")
```

# Comments

- If constant proportion of winters of each type over time, red, black and blue traces should each follow their line.
- Mixed (black) trace mostly does.
- Snowy (blue) trace mostly above its line: more snowy winters in past.
- Rainy (red) trace mostly below its line: fewer rainy winters in past.
- Formal test not known (by me), but devise one:
  - Measure observed dissimilarity between traces and lines (eg. area between them)
  - Simulate winters under null hypothesis that proportions not changing
  - Calculate dissimilarities for simulated winters
  - If observed dissimilarities unusually large compared to simulated, reject null and declare that proportions *not* constant.

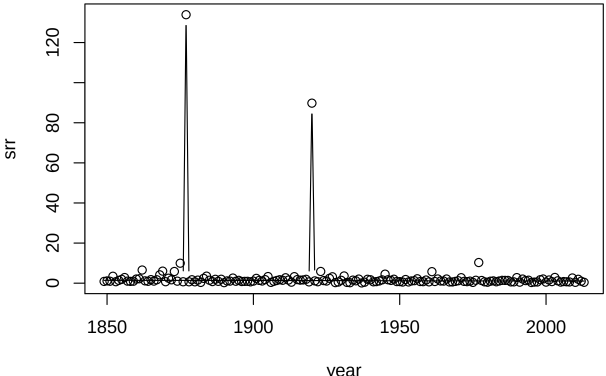
Section 6

# Snow-rain ratios

# Plotting against time

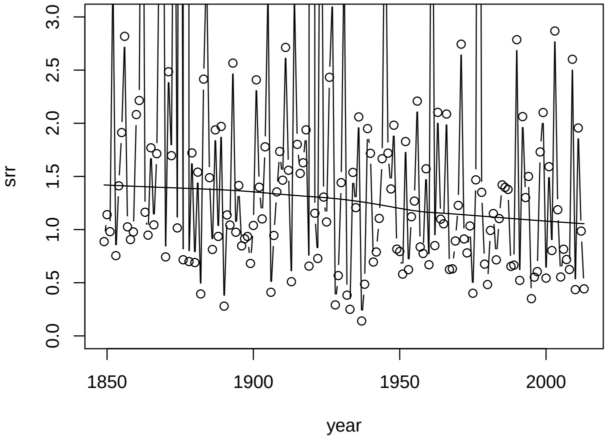
Calculate and attempt to plot against time:

```
> srr=snow/rain
> winter2$srr=srr
> plot(srr~year,type="b")
> lines(lowess(srr~year))
```



Some ratios extremely large, so truncate vertical scale:

```
> plot(srr~year,type="b",ylim=c(0,3))
> lines(lowess(srr~year))
```



# Testing for trend

- Small downward trend, apparently consistent amid noise.
- Haul out Mann-Kendall again:

```
> library(Kendall)
> res.1=MannKendall(srr)
> res.1
```

```
tau = -0.146, 2-sided pvalue =0.0054324
```

- Make (approximate) CI for Mann-Kendall correlation (up and down twice SE):

```
> mkci=function(res)
+ {
+       sds=sqrt(res$varS)
+       ci=c(res$S-2*sds,res$S+2*sds)/res$D
+       ci
+ }
```

- For this example:

```
> mkci(res.1)
[1] -0.25079964 -0.04099637
```

# Theil-Sen

- Downward trend real. How big? Theil-Sen slope:

```
> library(zyp)
> zyp.sen(srr~year,data=winter2)$coefficients

   Intercept          year
 7.520284410 -0.003247676
```

- Decrease about 0.003/year, about 0.52 over the 160 years. In line with lowess.
- Mann-Kendall and Theil-Sen results should be ok: have some serious outliers, but Mann-Kendall, Theil-Sen are resistant to them, and linear trend looks reasonable.

Section 7

# Vineland Station

# Why?

- Somewhere rural, not too far from Toronto
- On Niagara escarpment.
- Compare trends with Toronto: are Toronto's trends global warming, or is Toronto an urban heat island?
- Data 1930–2007.

```
> vineland=read.csv("Vineland - Control data.csv",header=T)
> head(vineland)

  Year     Dec.s.r    Jan.s.r  Type Total.snow Total.rain
1 1930 2.214285714 0.31786217 mixed       83.8      128.0
2 1931 1.485915493 3.50561798 snowy       61.2       43.4
3 1932 0.001766784 0.03573931 rainy       17.4      228.8
4 1933 0.163716814 0.06983240 rainy       20.8      120.1
5 1934 0.587500000 0.23293173 rainy       86.6       89.0
6 1935 1.281250000 0.30825688 mixed       76.0      102.3
```

Section 8

# Vineland Station winter type

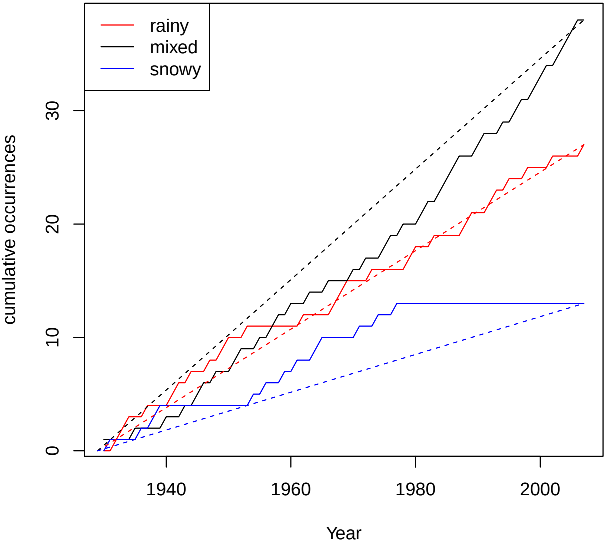
# Compute cumulative winters of each type

```
> attach(vineland)
> is.rainy=(Type=="rainy")
> rainy.sofar=cumsum(is.rainy)
> is.mixed=Type=="mixed"
> mixed.sofar=cumsum(is.mixed)
> is.snowy=Type=="snowy"
> snowy.sofar=cumsum(is.snowy)
```

# Checking

```
> head(data.frame(Year,Type,rainy.sofar,mixed.sofar,
+                 snowy.sofar),n=12)
   Year  Type rainy.sofar mixed.sofar snowy.sofar
1  1930 mixed           0           1           0
2  1931 snowy           0           1           1
3  1932 rainy           1           1           1
4  1933 rainy           2           1           1
5  1934 rainy           3           1           1
6  1935 mixed           3           2           1
7  1936 snowy           3           2           2
8  1937 rainy           4           2           2
9  1938 snowy           4           2           3
10 1939 snowy           4           2           4
11 1940 mixed           4           3           4
12 1941 rainy           5           3           4
```

# Rate evolution graph



# Comments

- In the past, fewer mixed years, more snowy years
- Trend to more mixed, fewer snowy (no snowy since 1977!)
- Similar picture to Toronto.

# The graph code

```
> plot(mixed.sofar~Year,type="n",ylab="cumulative occurrenc
> lines(rainy.sofar~Year,col="red")
> lines(mixed.sofar~Year,col="black")
> lines(snowy.sofar~Year,col="blue")
> legend("topleft",c("rainy","mixed","snowy"),col=c("red",'
> year1=min(Year)-1
> year2=max(Year)
> lines(c(year1,year2),c(0,max(rainy.sofar)),col="red",lty=
> lines(c(year1,year2),c(0,max(mixed.sofar)),col="black",lt
> lines(c(year1,year2),c(0,max(snowy.sofar)),col="blue",lty
```

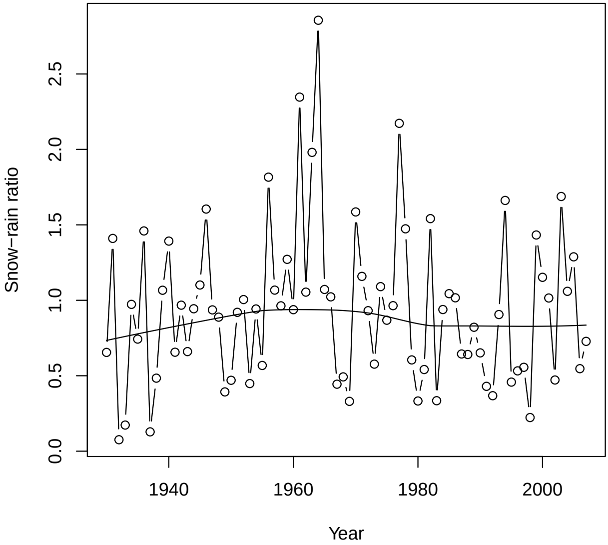
Section 9

# Vineland Station: snow-rain ratio

- I don't think I have Toronto snow-rain ratios for January and February separately, so calculated for whole winter each year:

```
> srr=Total.snow/Total.rain
> vineland$srr=srr
> plot(srr~Year,type="b",ylab="Snow-rain ratio")
> lines(lowess(srr~Year))
```

# Snow-rain ratio plot



# Comments

- Appears to be no trend:

```
> res.2=MannKendall(srr)
> res.2
tau = -0.00899, 2-sided pvalue =0.91068
> mkci(res.2)
[1] -0.1633543  0.1453723
```

No evidence whatever.

Section 10

# Comparing Toronto over same years

# Extracting the data

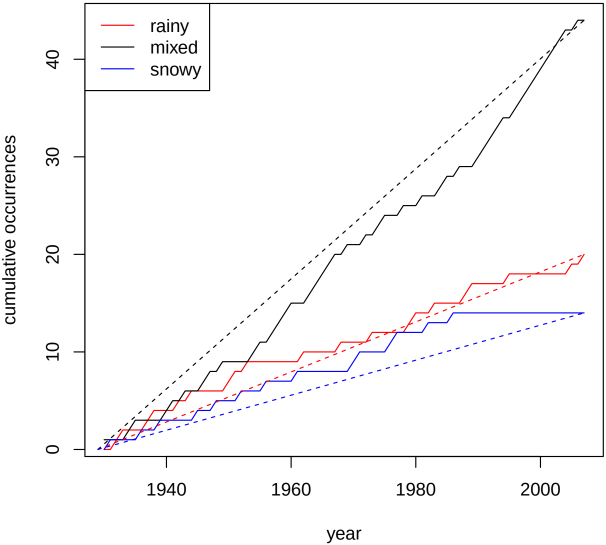
- `winter2` had Toronto data sorted by years. Select wanted years:

```
> attach(winter2)
> toronto=winter2[year>=1930 & year<=2007,]
> detach(winter2)
> head(toronto,n=5)
    year  snow  rain winter.type       srr
57  1930 118.8  82.4       mixed 1.4417476
90  1931  97.8  27.1       snowy 3.6088561
106 1932  88.9 231.8       rainy 0.3835203
165 1933  28.3 113.1       rainy 0.2502210
121 1934  82.9  53.9       mixed 1.5380334
> tail(toronto,n=5)
    year  snow  rain winter.type       srr
118 2003  84.6  29.5       mixed 2.8677966
52  2004 120.7 101.7       mixed 1.1868240
123 2005  82.0 147.9       rainy 0.5544287
129 2006  77.4  94.9       mixed 0.8155954
110 2007  87.9 122.3       rainy 0.7187244
```

# Cumulative winters of each type

```
> attach(toronto)
> is.rainy=(winter.type=="rainy")
> rainy.sofar=cumsum(is.rainy)
> is.mixed=winter.type=="mixed"
> mixed.sofar=cumsum(is.mixed)
> is.snowy=winter.type=="snowy"
> snowy.sofar=cumsum(is.snowy)
```

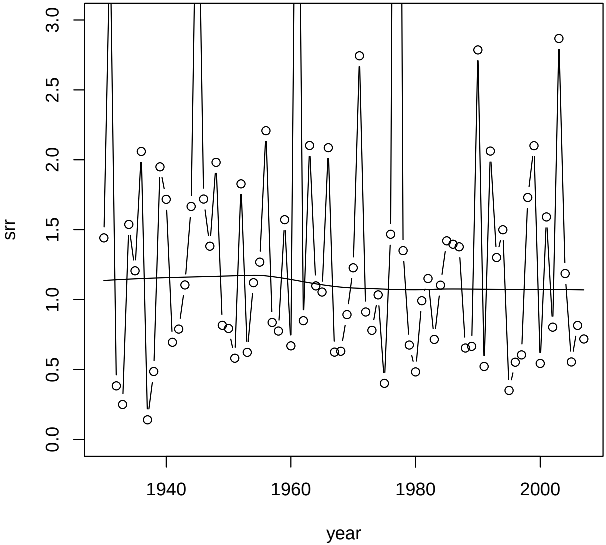
# Rate evolution graph



# Comments

- Toronto rate evolution graph looks a lot like Vineland Station's, over same time period:
  - a sharp decrease in snowy winters
  - an increase in mixed winters.

# Snow-rain ratio



# Comments

- No longer appears to be much of a trend:

  > *res.3=MannKendall(srr)*

  > *res.3*

  ```
  tau = -0.0523, 2-sided pvalue =0.50091
  ```

  > *mkci(res.3)*

  ```
  [1] -0.2066443  0.1020822
  ```

- This confuses me, since the time plot of the full data showed a steady downward trend, with the sharpest decline between about 1940 and 1950. May be because of using less data here.
- Toronto and Vineland Station appear to be showing similar trends since 1930. Both are showing a decrease in the proportion of snowy winters, and an increase in the proportion of mixed winters. Neither location appears to be showing any trend in snow-rain ratios in that period.