Interpreting and Smoothing Time Series

Repeated measurements of the *same variable* over a long period of *time* is called a *time series*. Examples include:

1. The number of U.S. exports per month for the last 60 months
2. The number of traffic fatalities each year for the past 50 years
3. Measurement of a person's body temperature every hour for 24 hours

Different from measuring, say body temperature, just once on each of 24 people!

We will look at

- **Graphing time series.** We will use a kind of scatter plot called a *time series plot* (also called a *line plot*.

- **Interpreting time series.** What are the important features to look for in a time series plot? *Why* are they there?

- **Smoothing time series.** We can "smooth out" random variability in the time series to see what the underlying features are.

Interpreting Time Series

Random variation gives time series lots of small, short-term fluctuations. We want to the the overall pattern, and ignore the small fluctuations:

- *Trend:* Is there a linear trend in the data (increasing or decreasing?)? Is there a long-term, curvilinear trend (e.g. down at first, up in the middle, down at the end)? *Why: What could cause these patterns?*

- *Cycles:* A pattern that repeats every 3 minutes, every hour, every 2 days, every month, every quarter, every year, etc. (or smaller or larger units of time). *Why* is that the cycle?

- *Seasonal Variation:* In economic time series, yearly cycles are called seasonal variations. *Why* are they there?

- *Small Scale Fluctuations:* Smaller but noticable deviations from the pattern established by the trend and the cycles or seasonal behavior. *Why* are they there?

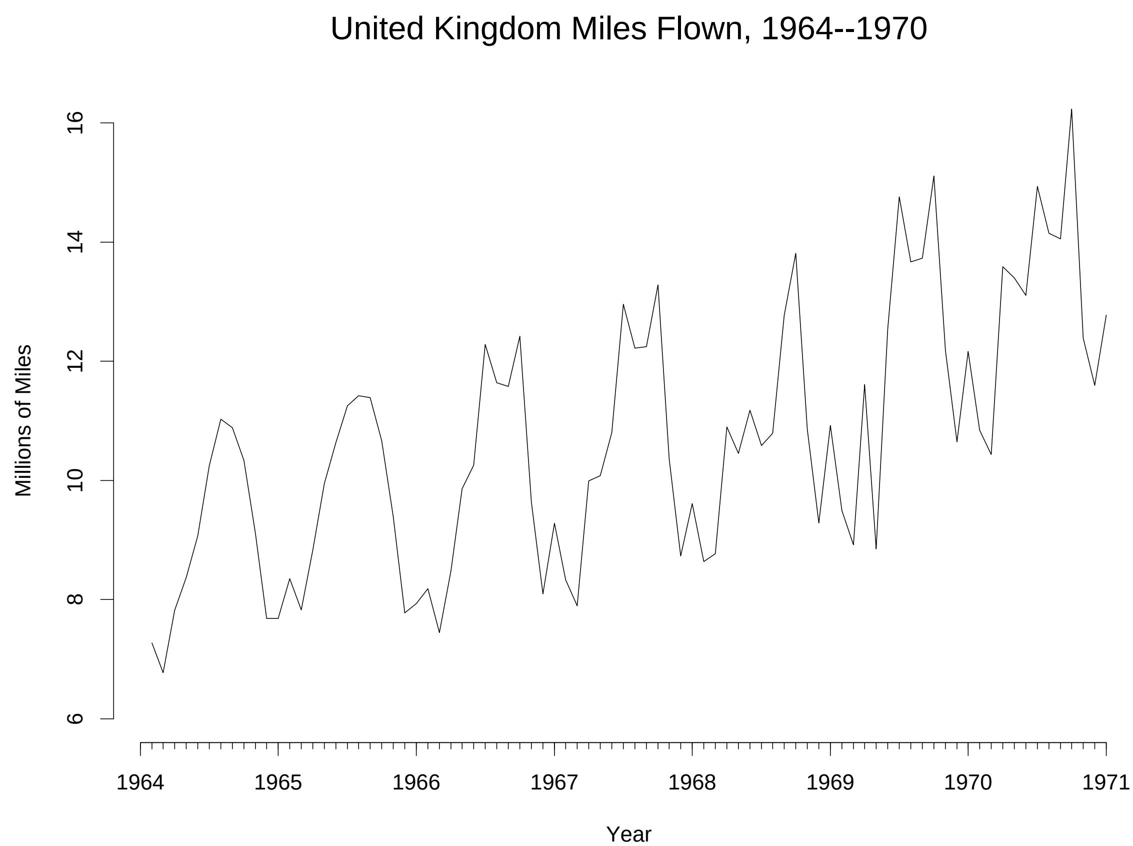
- *Changing Variability:* When the time series has cycles, does the peak-to-valley distance change over time? *Why* does the variability change?

Case Study: Airline Miles Flown

The table below shows the monthly total airline miles flown in the United Kingdom for the seven years 1964–1970.

United Kingdom Miles Flown, 1964–1970, in Millions of Miles

Month	1964	1965	1966	1967	1968	1969	1970
Jan.	7.269	8.350	8.186	8.334	8.369	9.491	10.840
Feb.	6.775	7.829	7.444	7.899	8.772	8.919	10.436
Mar.	7.819	8.829	8.484	9.994	10.894	11.607	13.589
Apr.	8.371	9.948	9.864	10.078	10.455	8.852	13.402
May	9.069	10.638	10.252	10.801	11.179	12.537	13.103
June	10.248	11.253	12.282	12.953	10.588	14.759	14.933
July	11.030	11.424	11.637	12.222	10.794	13.667	14.147
Aug.	10.882	11.391	11.577	12.246	12.770	13.731	14.057
Sept.	10.333	10.665	12.417	13.281	13.812	15.110	16.234
Oct.	9.109	9.396	9.637	10.366	10.857	12.185	12.389
Nov.	7.685	7.775	8.094	8.730	9.290	10.546	11.594
Dec.	7.682	7.933	9.280	9.614	10.925	12.161	12.772



<u>Features of a times series</u>

You can see the following features in the time series:

- *<u>Trend</u>*
- *<u>Cycles</u>*
- *<u>Seasonal Variation</u>*
- *<u>Small Scale Fluctuations</u>*
- *<u>Changing Variability</u>*

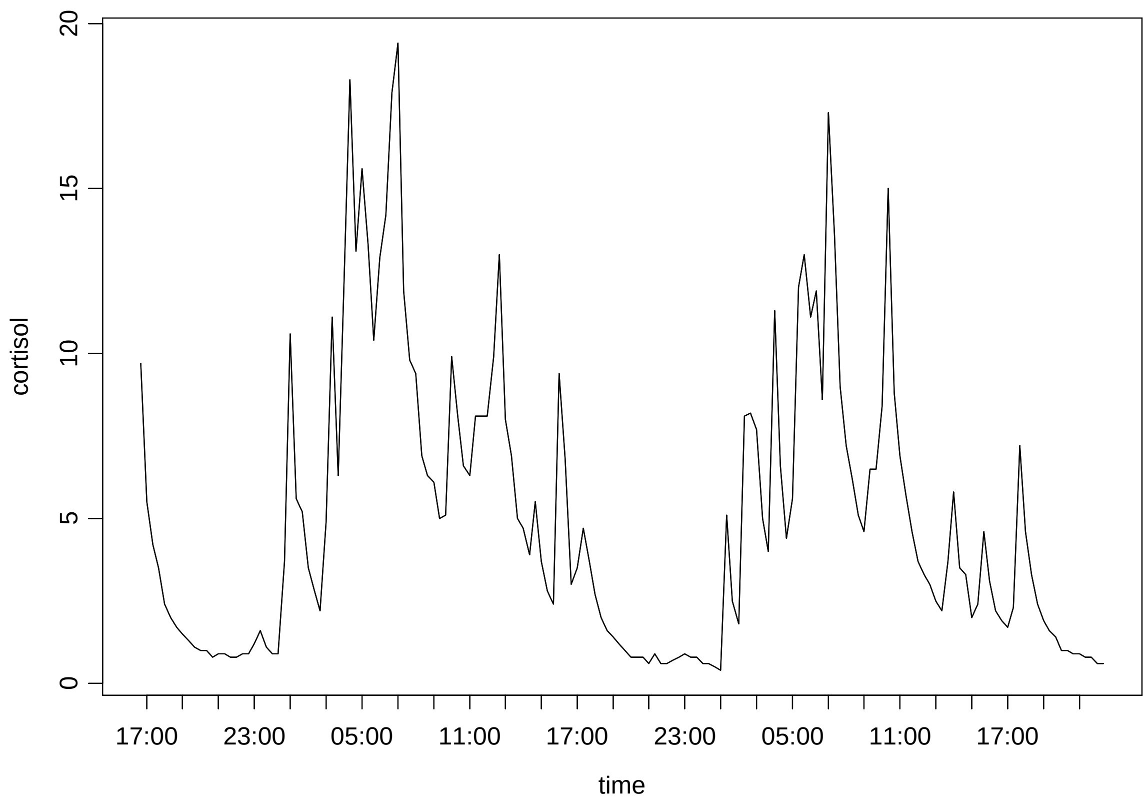
Smoothing Time Series

- We want to see the "pattern" and not the small-scale random fluctuations. We can do this with *smoothing*.
- *moving average*: Replace each observation with an average made up of 3 values: the observation before it, the observation after it, and the observation itself.

Case Study: Cortisol levels throughout the day

In this case study we will explore the scatter plot of a time series of blood plasma cortisol levels (the response variable) obtained every 20 minutes for 54 consecutive hours starting at 16:40 (4:40 p.m.) from a single healthy volunteer. This allows us to study the *circadian (daily) rhythms* of humans.

The time series plot is displayed on the next page.



The time series plot shows cyclic up-and-down behavior:

- Lows around 10:20pm on the first day, 10:30pm the second day, and 10:20pm the third day.
- Highs are around time 5:40am on the first day, and 6:20am on the second day.

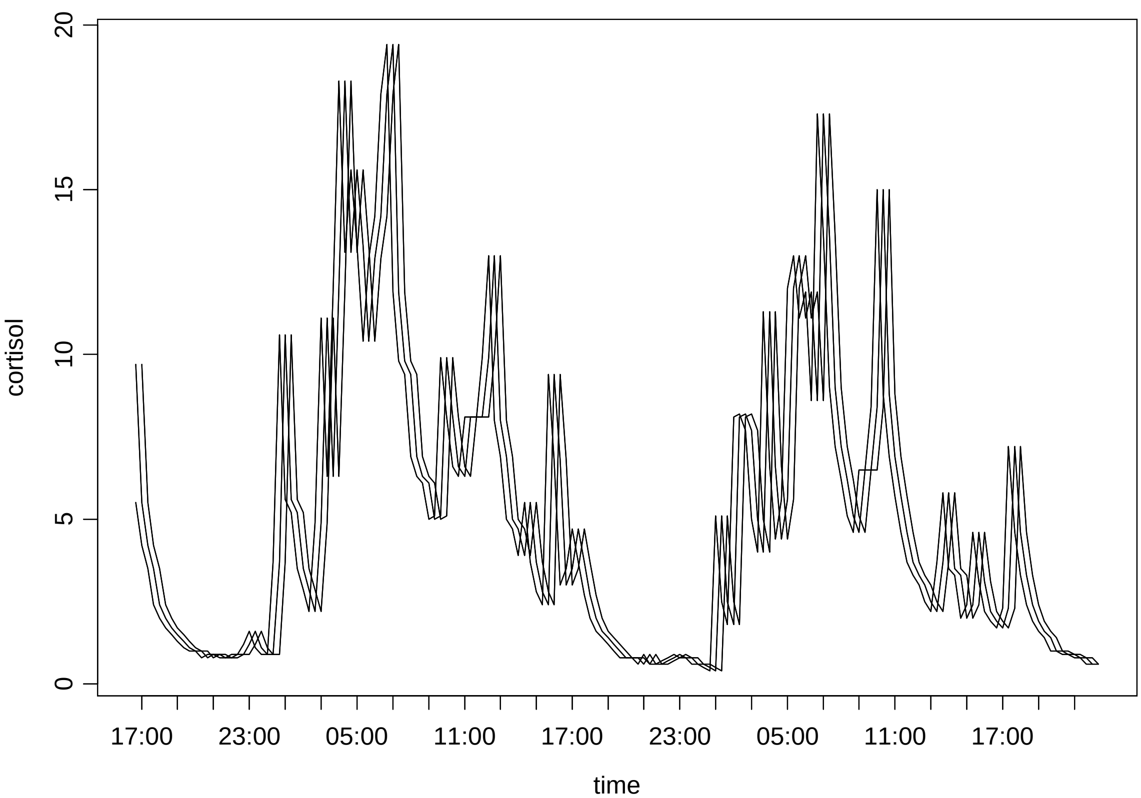
Which of the smaller variations are important?

Smoothing: Moving average

Replot the series three times:

- once in its original position,
- once with every observation shifted to the *left* one place,
- and once with every observation shifted to the *right* one place,

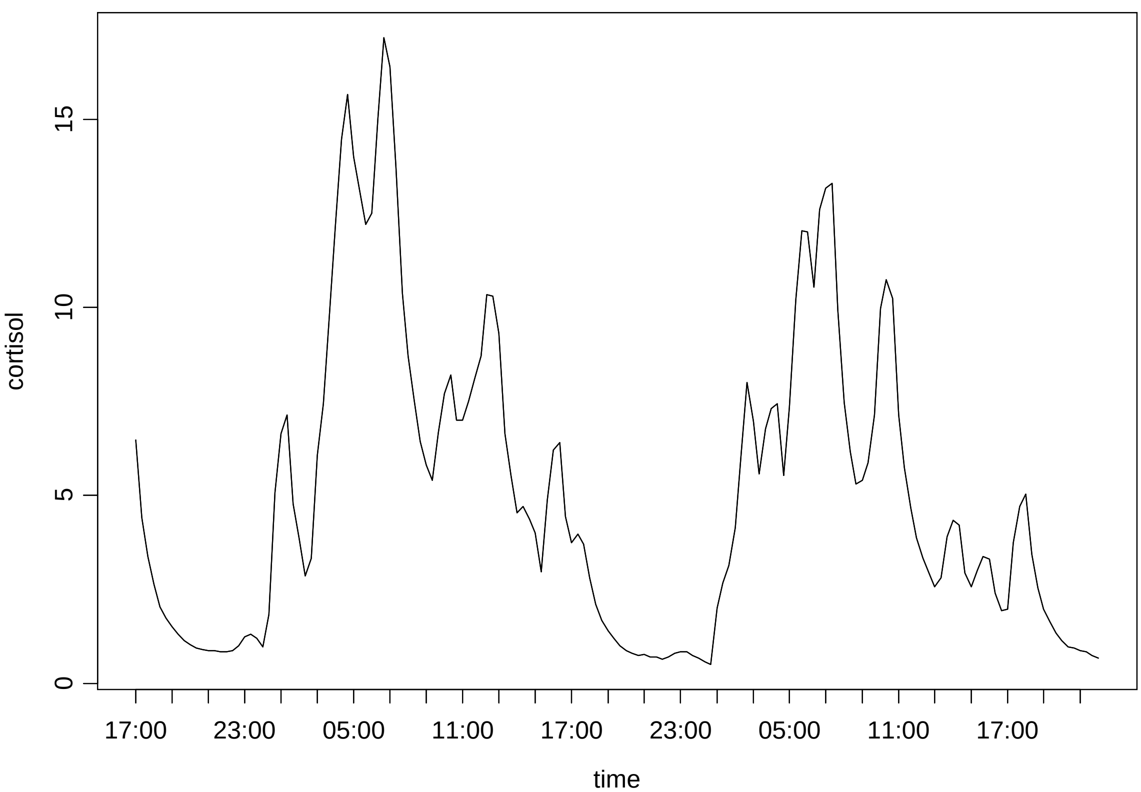
and then average them.



This probably makes your eyes crazy.

Now average together the three observations above each time point,

$$\textit{(smoothed obs)} = \frac{\textit{(obs to the left)} + \textit{(true obs)} + \textit{(obs to the right)}}{3}$$



Interpretation

- Minor variations have been smoothed away
- Lows around 10:30pm and highs around 6:30am are clearer.
- Even after smoothing, some smaller fluctuations are visible.

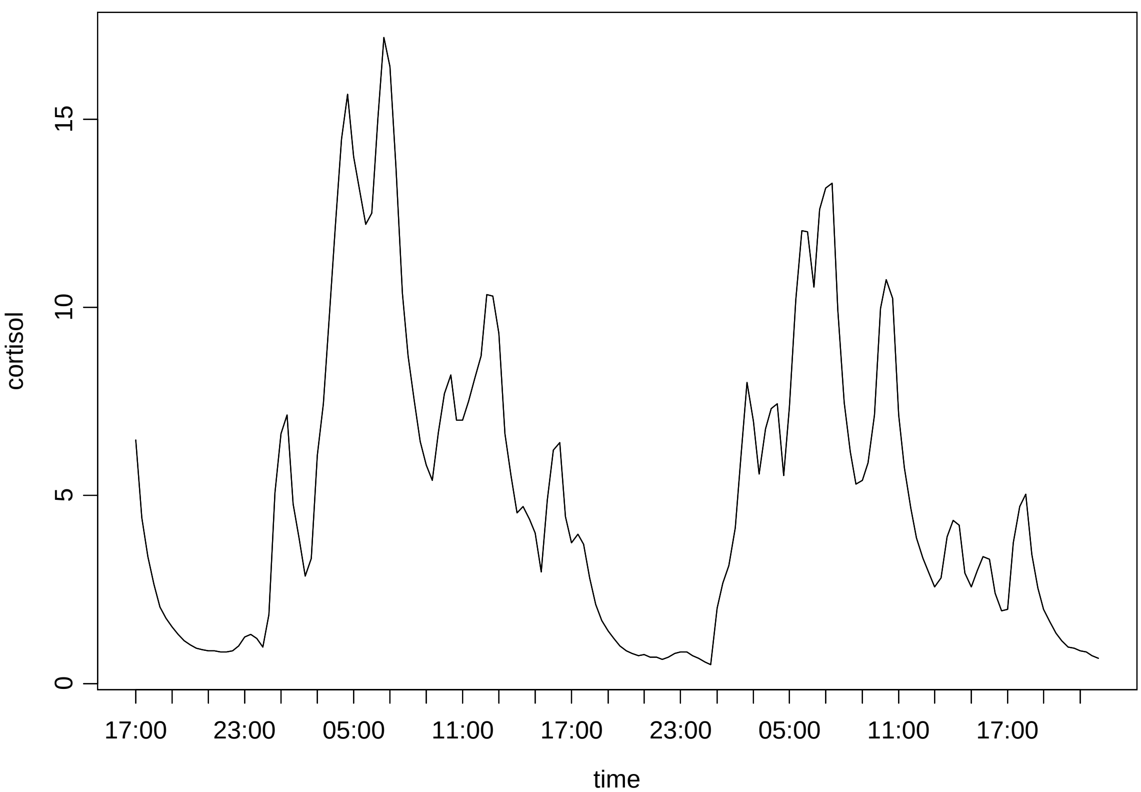
Sometimes the “simple” moving average is too smooth.

We can “rough it up” a bit, by double-counting the “real” time series.

This is called a *weighted moving average*.

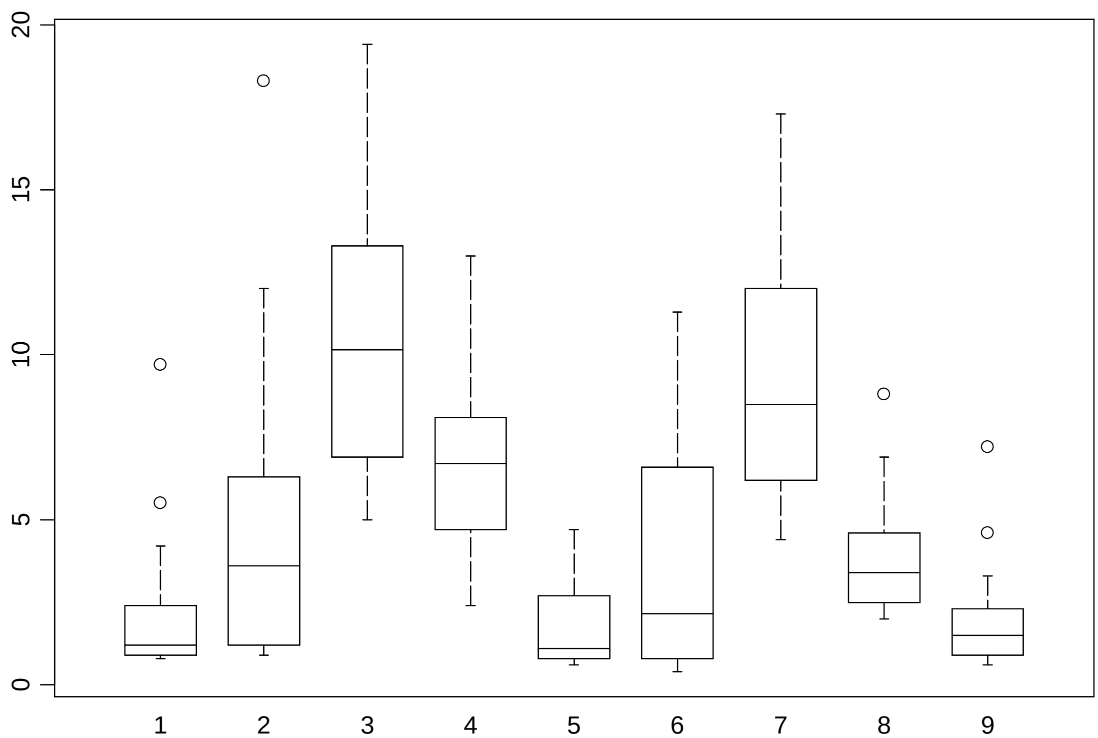
(smoothed obs)

$$= \frac{\text{(obs to the left)} + 2 \times \text{(true obs)} + \text{(obs to the right)}}{4}$$



Earlier in the semester we looked at “median trace plots” as another way to smooth out a scatter plot to see the relationship between the variables.

Here is a median trace plot of the cortisol data. After some experimentation, I divided the 162 measurements up into 9 groups of 18 measurements (3 hours) each.



In general a moving average smooth preserves more detail than the “median fit smooths” that we did for other scatter plots.

However, the boxplots make it *very clear* that there is changing variability in this data—the variability is much lower near the valleys than near the peaks.

If we wanted the variability to look more equal across the time series, we could *transform* the data in some way.

For example, *logarithms* squash large values much more than small ones.