Todd Garner

DS6306 Week 11 FLS

"Forecasting is difficult. Especially about things in the future" Yogi Berra

Question 1 – Activity 1

- The first thing I noticed (only running a piece of the chunk) is that the units are not immediately apparent.
- The function SES (since I've not seen this one yet):

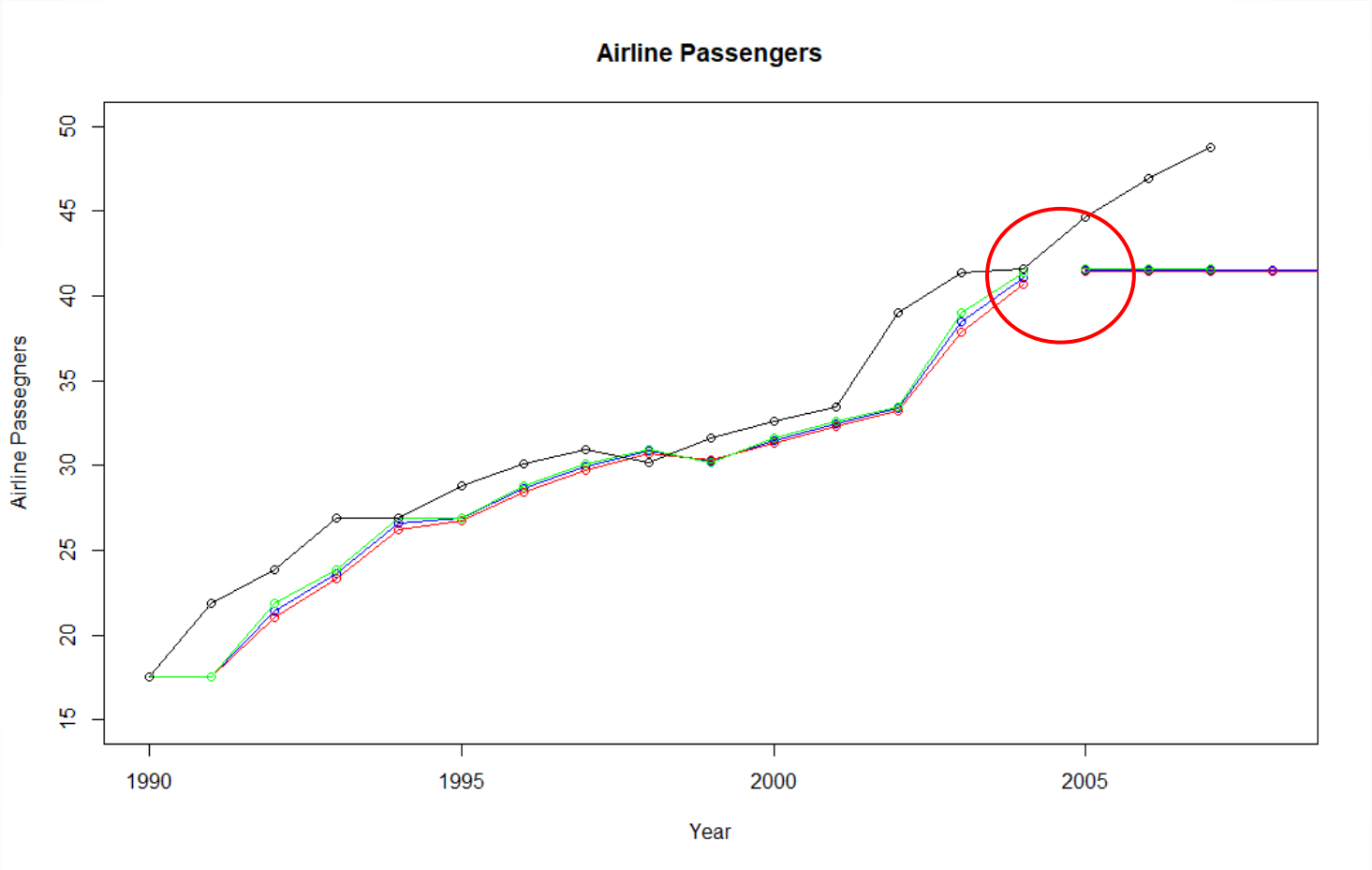
ses(	
y,	numeric vector or time series
h = 10,	number of periods
level = c(80, 95),	confidence level for prediction intervals
fan = FALSE,	setting for "fan plots"
initial = c("optimal", "simple"),	smoothing settings for parameters
alpha = NULL,	value of smoothing parameters
lambda = NULL,	Box-Cox transformation parameter, NULL = transformation ignored
biasadj = FALSE,	back transformation settings – either median or mean
x = y,	Deprecated, included for backwards compatibility (perhaps in FLS I can learn what that means?)
...	other arguments passed to the forecast
)	

Activity 1 - continued

- The models, in general, use weighting to emphasize nearer term items and de-emphasize items that are away from the target timeframe.
- Three different "fits" are produced.
  - The differences are centered around the smoothing parameters
- The measures calculated are:
  - ME: Mean Error
  - RMSE: Root Mean Squared Error
  - MAE: Mean Absolute Error
  - MPE: Mean Percentage Error
  - MAPE: Mean Absolute Percentage Error
  - MASE: Mean Absolute Scaled Error
  - ACF1: Autocorrelation of errors at lag 1.

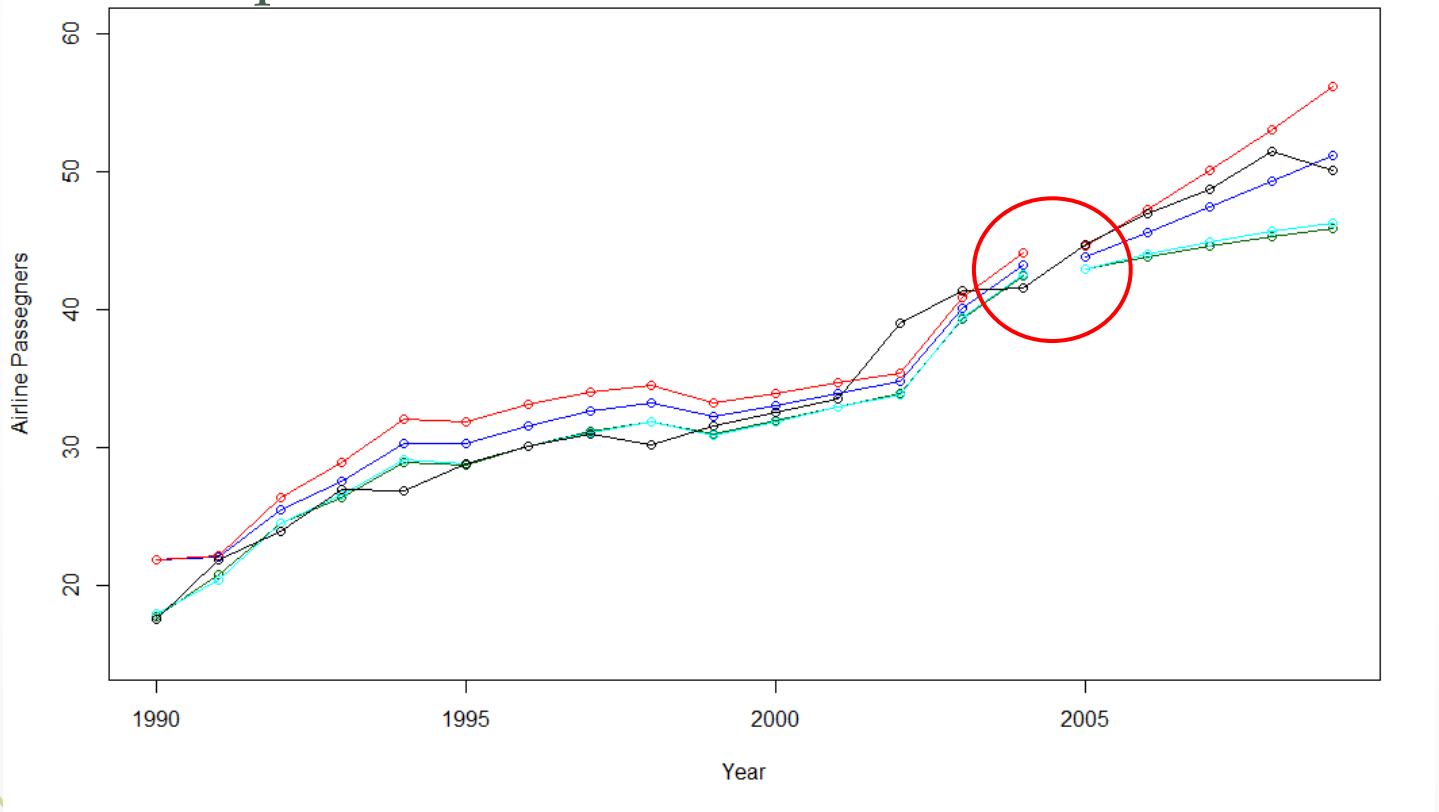
- Reading numerous articles on stackoverflow.com and others, these measures imply various measures to be chosen "best" by the particular application answer you're seeking.
- By "fiddling" with the parameters, I was able to test varying sensitivities and learned much about exponential smoothing functions.
- The closer alpha was to 1, the more the lines converged.
- No real seasonality was noticed in the data, even though it's over 39 years. As there is just data point per year, this makes sense. It just appears that we have an upward trending line over time. In all, not the most useful data set.
- No decomposition is available with only one data point per year.
- In the different ses and accuracy functions, we appear to have dropped out some data between 2004 & 2005. I'm not sure what caused this.

```
fit1 = ses(air, initial = "simple",alpha = .9,h = 10)
fit2 = ses(air,initial = "simple",alpha = .8, h = 12)
fit3 = ses(air, h = 3)
```



Activity Two – holt(function)

- With the given code, the resulting chart is, again with a line missing. If I search the data set, I see that these particular observations are very close in value (41.3863, 41.59655). Still a head scratcher. Why would the line be missing if the data point is there?



Activity Two - continued

- Again, we see this hole in the chart (previous slide). Why don't they connect?
- The Holt-Winters function is widely used and has some inherent capabilities that we'll take advantage of and then we have the ability to "tune" them manually. Let's dig in. Again, using the same data set, we only have one data point per year so tuning may be an illusion.
- These are the tuning parameters:
  - **alpha**: the "base value". Higher alpha puts more weight on the most recent observations.
  - **beta**: the "trend value". Higher beta means the trend slope is more dependent on recent trend slopes.
  - **gamma**: the "seasonal component". Higher gamma puts more weighting on the most recent seasonal cycles.
  - The accuracy functions have produced these tables of data. I need to interpret the data given these attributes.

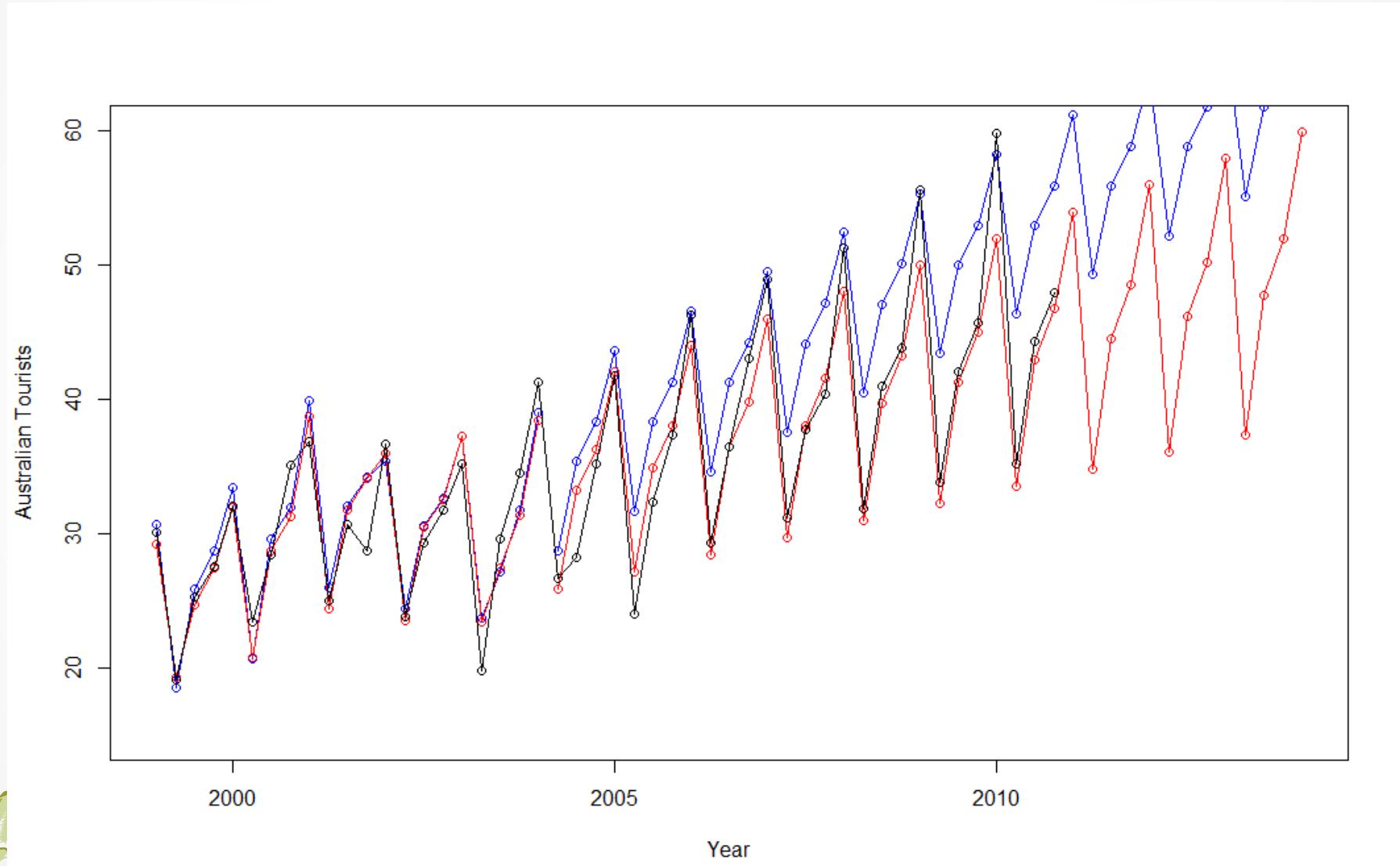
- From the chart on the previous page, we can see that the lines generally track but "fan out" toward the end.
- There seems to be some "tussles" about which metric to utilize and MAE uses median and RMSE uses the mean, but appear to come out on top of the others. Your application and answer you seek dominate which to use.
- In the table ahead, let's focus on those two and see what differences they deliver.

Accuracy functions returned the following:

```
> accuracy(fit1h, airTest)
                    ME     RMSE      MAE       MPE     MAPE      MASE       ACF1 Theil's U
Training set -1.0292272 2.202869 1.772637 -4.485612 6.364749 0.9671310  0.2088940        NA
Test set      0.9405879 1.444776 1.371076  1.947628 2.808140 0.7480438 -0.3313068 0.7212626
> accuracy(fit2h, airTest)
                    ME     RMSE      MAE       MPE     MAPE     MASE      ACF1 Theil's U
Training set -2.010092 2.908422 2.565680 -7.724285 9.137707 1.399806 0.3109081        NA
Test set     -1.836519 2.883708 1.859959 -3.672580 3.725068 1.014773 0.1029680  1.403487
> accuracy(fit3h, airTest)
                    ME     RMSE      MAE       MPE     MAPE     MASE      ACF1 Theil's U
Training set 0.3317307 1.653933 1.084962 0.7156323 3.371681 0.591943 0.1675855        NA
Test set     3.8715783 4.131618 3.871578 7.8802382 7.880238 2.112290 0.2453622  2.086233
```

- RMSE delivers in the same units of the observations. So, in fit2h, the RMSE training set and test set are: 2.908 train, 2.8837 test. In our summary, the high and low are 51.488 and .319, respectively. So, not "too" bad, but we'd always like better. 0 is virtually impossible to achieve.
- MAE also delivers in the same units as observation but utilizes the median versus the mean in RMSE. Overall, MAE provides better results.

Activity 3 - Seasonality



Activity 3 - continued

- Straight out of the box, we can see the effects of seasonality as we are given quarterly data. It is a measure of tourists visiting Australia. The definition: *Quarterly visitor nights spent by international tourists to Australia. 1999-2010.* This definition is a bit confusing, so I dug further.
- This would appear to be the number of visits (arrivals) of non-resident persons from other countries without regard to the length or purpose of their stay on a quarterly basis, by quarter. The data is in thousands. This is a presumption given my research on Australia's tourist web site. https://www.abs.gov.au/statistics/industry/tourism-and-transport/overseas-arrivals-and-departures-australia
- This would account for the seasonality as well as the growing numbers by year. This data appears to be much better modeled using the HW function:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-0.4377946	2.266717	1.882623	-1.842903	6.492401	0.6090876	-0.04231276	NA
Test set	-5.0401911	6.019319	5.172738	-14.174410	14.398562	1.6735428	-0.13264014	0.5996539

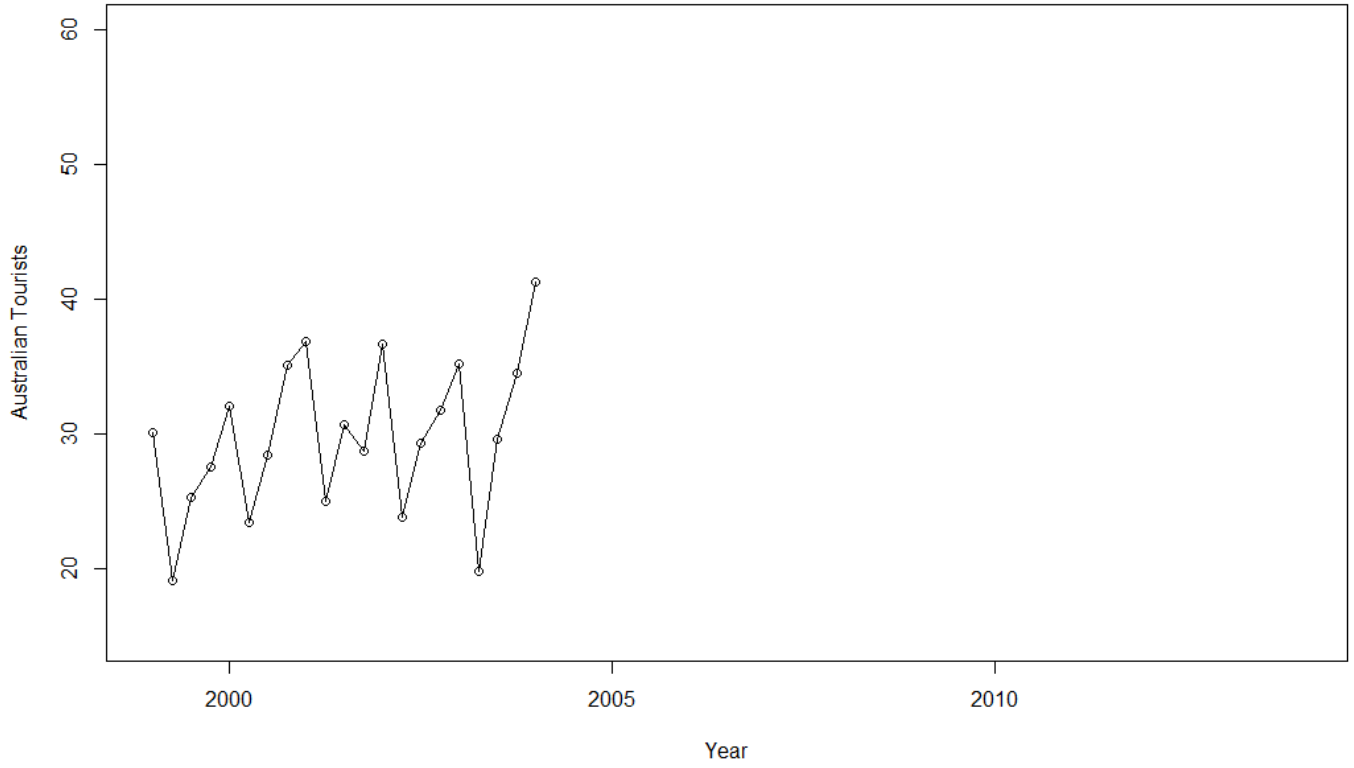
> accuracy(fit2s,austourists)

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	0.07223675	2.180841	1.626375	-0.1303281	5.531772	0.5261833	0.09930429	NA
Test set	0.88033413	2.627779	1.933411	1.3152869	4.865291	0.6255191	0.26125025	0.2670124

- As the data is in thousands, these errors are also in the thousands. Again, not as close as we'd like but given the applicability, they could be acceptable.

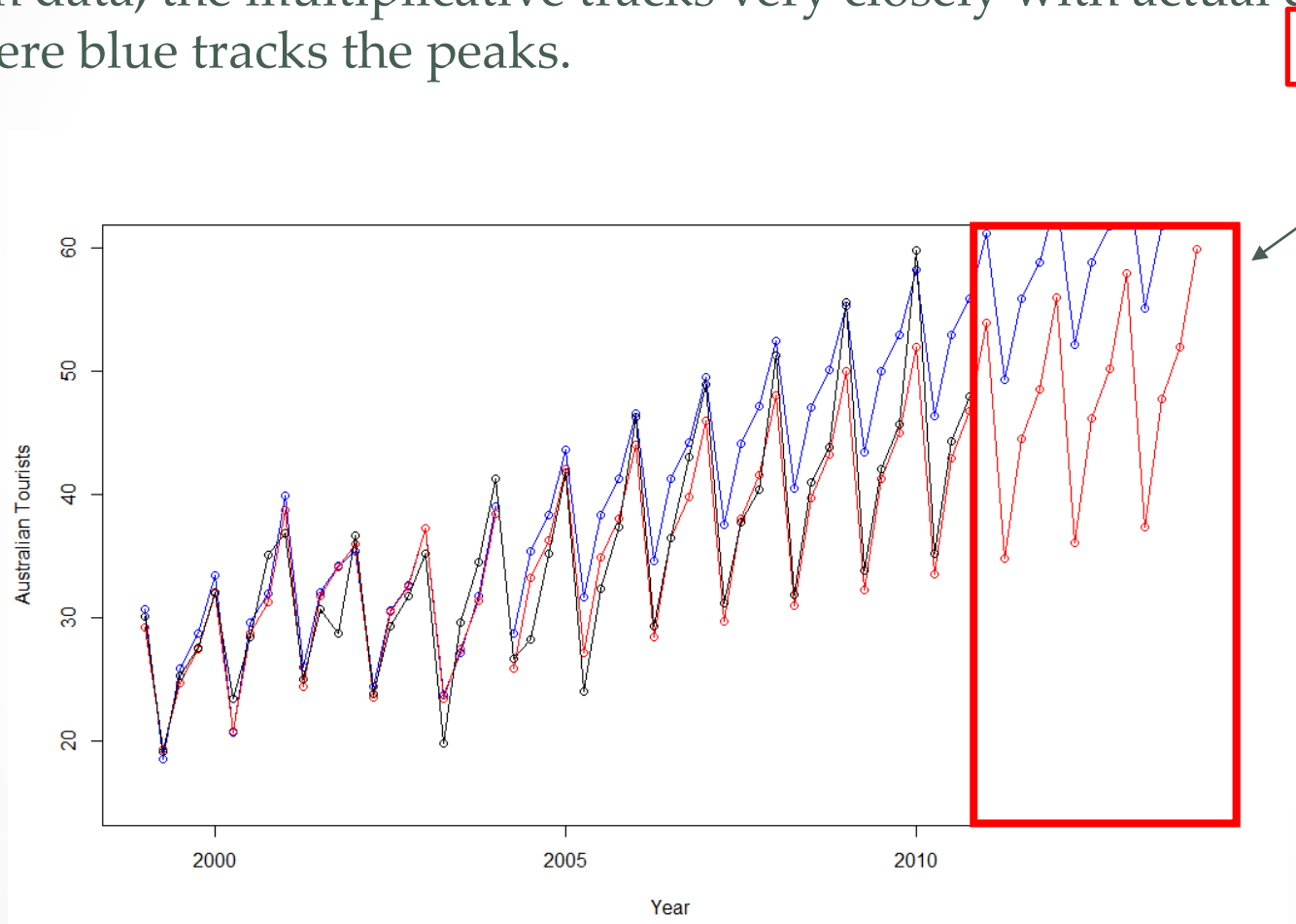
Activity 3 - continued

- This is a chart of the base data. The chart on the next page shows "additive" and "multiplicative" smoothing and prediction.



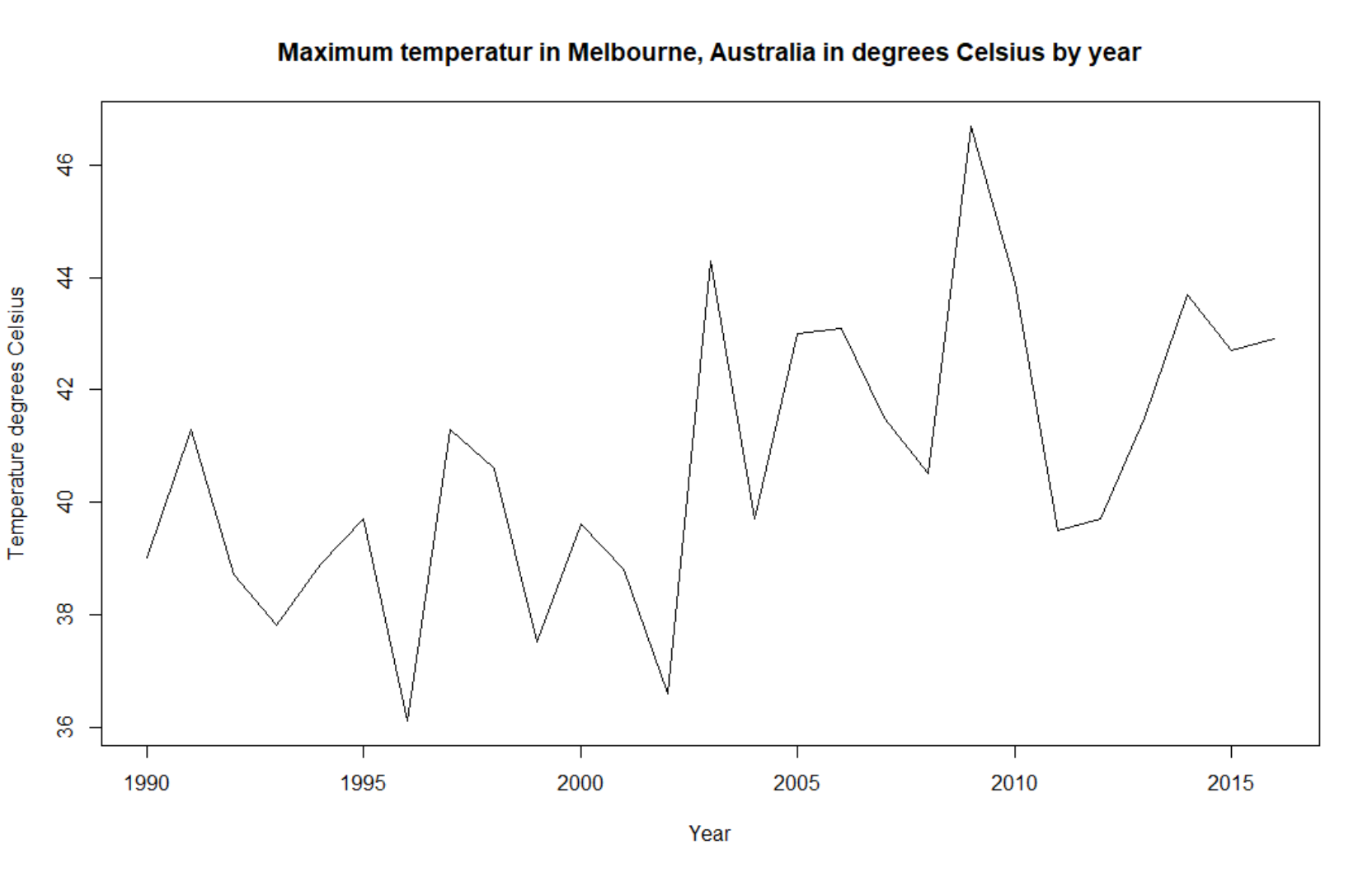
Activity 3 - continued

- Blue is “additive” and the red lines are “multiplicative.” in the region of known data, the known data, the multiplicative tracks very closely with actual data, except for the peaks where blue tracks the peaks.



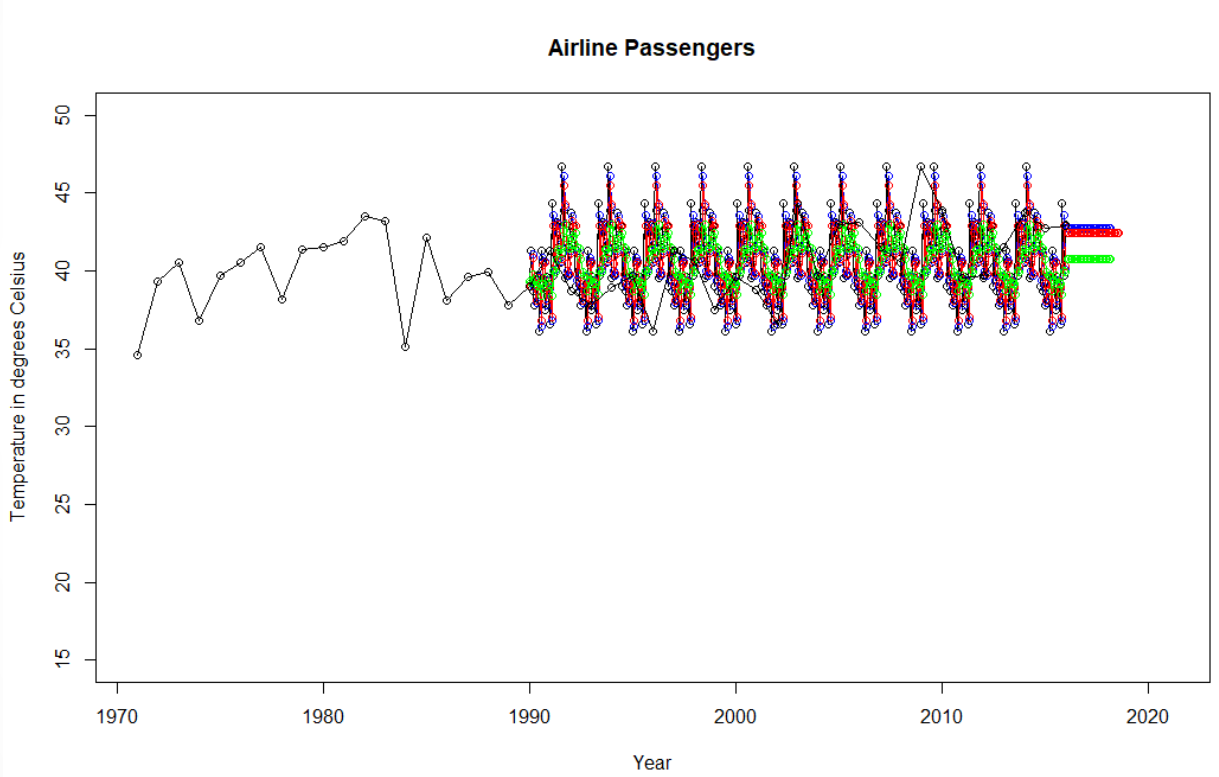
Activity 4

Maximum annual temperatures (degrees Celsius) for Moorabbin Airport, Melbourne.
1990(modified)-2016



Activity 4 - continued

- I've obviously got some problems in my time series and fit. Worked on this for hours and ran out of time.



Questions and Takeaways

- What does “deprecated” mean and “backward compatibility?
- I’m struggling with the predictive functions in time series. Can we discuss this in Live Session?
  - Which ones to use?
  - How to determine which is most appropriate?
  - How to set up the dates in the various ancillary functions?
- I’m also struggling with this error: **“Error in window.default(x, ...) : 'start' cannot be after 'end’”**