

MATH550/SCC461: Statistics in Practice
Lab 4
Assessment
Tom Palmer

Deadline for coursework submission is **9am Monday 30th October 2017**.

Assessment

<http://www.bom.gov.au/australia/stormarchive/>

The Severe Storms Archive of the Australian bureau of Meteorology contains data relating to recorded Severe Thunderstorm and related events in Australia dating back to the 18th Century.

Data can be downloaded from: <http://www.bom.gov.au/australia/stormarchive/>, and is also on Moodle.

In the file `Australian_severe_storms_1975-2015.csv` all Australian storm events in a 40 year period from 28th September 1975 to 28th September 2015 are recorded.

The data consists of a row for each severe weather event and 14 columns. The last 6 columns consist of Comments which we will ignore whilst the first 8 columns are:

Variable(s)	Description
Event.ID	Unique event ID
Database	Type of severe storm event.
ID	A secondary ID variable.
Date.Time	Date and time in the format dd/mm/yyyy hh:mm.
Nearest.town	Nearest town to the event.
State	Australian state in which the event was recorded.
Latitude	Latitude of the event.
Longitude	Longitude of the event.

Australia has three times zones. Australian Eastern Standard Time (AEST), Australian Central Standard Time (ACST) and Australian Western Standard Time (AWST). These time zones are 10 hours, 9.5 hours and 8 hours ahead of Coordinated Universal Time (UTC), respectively.

AEST applies to: New South Wales except Yancowinna County, which includes the city of Broken Hill; Victoria; Queensland; Tasmania; and the Australian Capital Territory. During daylight saving time, AEST becomes Australian Eastern Daylight Time (AEDT) and clocks are advanced to UTC+11. Note that Queensland does not observe daylight saving time.

ACST applies to: South Australia; the Northern Territory; and Yancowinna County, which includes the city of Broken Hill, in New



(a) A flooded bridge.



(b) 2-inch hail.



(c) Mud-slide on a road.



(d) An incoming storm.



(e) Big waves.



(f) Wooden plank stuck in a tyre.

Figure 1: Photographic evidence of storm events and their damaging effects.

South Wales. During daylight saving time, ACST becomes Australian Central Daylight Time (ACDT), and clocks are advanced to UTC +10:30. The Northern Territory does not observe daylight saving time. Finally, AWST applies to Western Australia.

Tasks

Download the data set, 'Australian_severe_storms_1975-2015.csv', from the 'Datasets' section on the MATH550 Moodle page.

Submit a compiled report that performs the following:

1. Set your working directory to the folder which contains the data.
2. Read in the data.

Print the dimensions of the Storm Events data frame.

[1 mark]

3. Clean the data by removing the last 6 columns associated with comments along with ID. Also remove Waterspout events from the database.

Print the first few rows of the cleaned Storm Events data frame along with the dimensions of the dataframe.

[1 mark]

4. Add a column to your dataframe containing the time zone of each event using the following `OlsonNames()` classifications.

Table 1: List of `OlsonNames()` classifications of time zones.

State	Abbreviation	Time Zone
Queensland	QLD	Australia/Queensland
New South Wales (excluding Broken Hills)	NSW	Australia/NSW
Broken Hills, New South Wales	NSW (Broken Hills)	Australia/Broken_Hill
Victoria	VIC	Australia/Victoria
South Australia	SA	Australia/South
West Australia	WA	Australia/West
Tasmania	TAS	Australia/Tasmania
Northern Territory	NT	Australia/North
Australian Capital Territory	ACT	Australia/ACT

Hint: Write a function to allocate a time zone based upon the input of State and `Nearest.town`. (Note `Nearest.town` is only relevant to New South Wales.) Use a for loop and the time zone allocation function to assign a time zone to each event.

[3 marks]

5. Parse the date, time and time zones from the necessary columns to create a new variable in the data frame which converts the time into UTC.

Print the first few rows of the Storm Events data frame.

[1 marks]

6. Create new variables for the month and year of each event.

Print the first few rows of the Storm Events data frame.

[1 marks]

7. After discarding Waterspout events there are five types of events left in the data; Rain, Hail, Lighting, Wind, and Tornado.

Create a new data frame which contains the total number of counts for each of the above type of events for each of the twelve months over the forty year period.

On a single plot, plot the total number of counts of each event against month.

[3 marks]