

Hydrology and Water Quality of the Mangatainoka Catchment

Prepared by David R. Maidment
Waterways Centre for Freshwater Research
University of Canterbury

and Brent Watson
Horizons Regional Council
Palmerston North

February 2018

Goals of the Exercise

This exercise introduces you to observations data on rainfall, flow and water quality measured in the Mangatainoka Catchment, and in particular on the Mangatainoka River at Pahiatua Town Bridge. Two types of data are used: *daily time series* that represent the total rainfall, or the average over the day of automatically recorded water flow and quality, and *instantaneously sampled data* that measure coliform bacteria obtained by laboratory analysis of water samples taken at various instantaneous time points during the past two summers. A regression model is used to relate the instantaneous E. Coli concentrations to contemporaneous flow and turbidity values, and this model is then coupled with daily values of those quantities to estimate E. Coli on a daily basis, and to compute the mean daily Coliform bacteria load in the Mangatainoka River at Pahiatua Town Bridge.

Computer and Data Requirements

To carry out this exercise, you need to have a computer, which runs Excel. You'll use data in a file **MangatainokaObservations.xlsx** that is in the Ex2 folder on the course web site, and is also accessible at: <http://www.cae.utexas.edu/prof/maidment/Canterbury/Ex2/MangatainokaObservations.zip>

Procedure

Instantaneous Observations Data

The first spreadsheet [E.coli Samples](#) contains instantaneous values of E Coli that were taken as bottled water samples particular points in time from Mangatainoka River at Town Bridge, and the values of other quantities that were simultaneously being automatically monitored at that location:

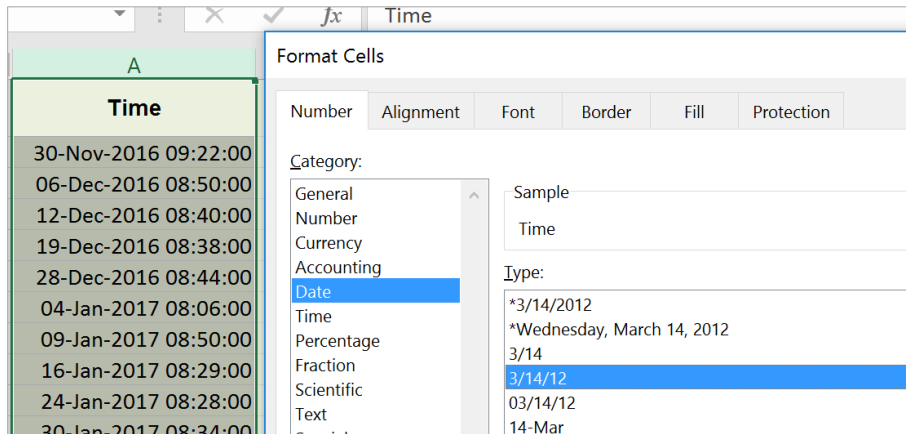
- E. Coli by MPN (most probable number) of Coliform Units (CFU) per 100 mL of water
- Flow or Discharge in liters per second,
- Water Temperature in °C,

- Dissolved Oxygen Saturation as a % (this measures for the given water temperature the % of the oxygen concentration that would have fully saturated the water),
- Turbidity (the degree to which light is scattered by particles suspended in a liquid) measured in FNU (Formazin Nephelometric Units)

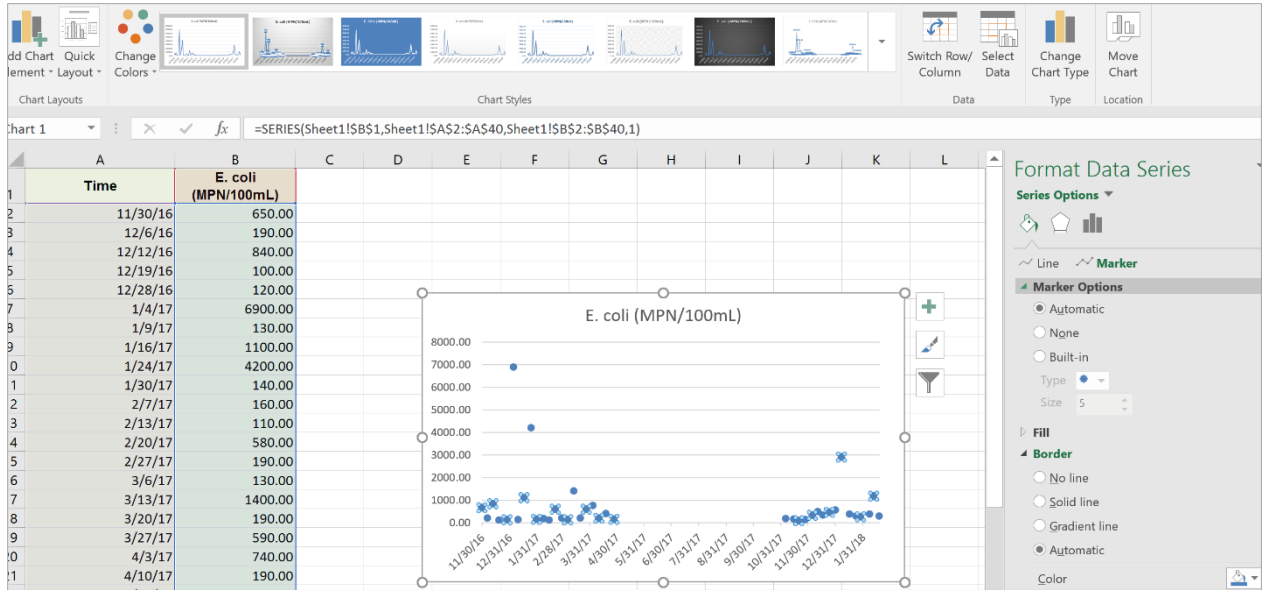
For example, on 30 November 2016, at 9.22AM, there was an E. Coli measurement of 650 CFU/100mL, when the flow was 32642 liters per second (or 32.642 m³/s), the water had a temperature of 13.1°C, the DO saturation was 91.91% and the Turbidity was 4.00 FNU. The value shown in blue on the right hand side is the daily mean flow for the same day. You can see that the value on 30 Nov 2016 was 90590 liters per second, which is quite different than the instantaneous value at 9:22AM in the morning. This means that there was a discharge hydrograph event going on at this time, and the flow was varying quite a lot during this day.

Site Name	Time	E. coli by MPN (HRC)	Flow (l/s)	Water Temperature [Dissolved Oxygen sensor] (°C)	DO Saturation (%)	Turbidity: Combined Point Sample (FNU)	Daily Mean Flow (l/s)
Mangatainoka at Pahiatua Town Bridge	30-Nov-2016 09:22:00	650.00	32642.00	13.10	91.91	4.00	90590.00
Mangatainoka at Pahiatua Town Bridge	06-Dec-2016 08:50:00	190.00	13540.00	15.40	95.16	1.51	13080.00
Mangatainoka at Pahiatua Town Bridge	12-Dec-2016 08:40:00	840.00	11156.00	15.40	94.94	2.09	10790.00
Mangatainoka at Pahiatua Town Bridge	19-Dec-2016 08:38:00	100.00	10518.00	15.20	91.55	1.24	21500.00
Mangatainoka at Pahiatua Town Bridge	28-Dec-2016 08:44:00	120.00	5354.00	15.90	99.93	0.65	5801.00
Mangatainoka at Pahiatua Town Bridge	04-Jan-2017 08:06:00	6900.00	49127.00	14.70	88.91	16.92	47090.00
Mangatainoka at Pahiatua Town Bridge	09-Jan-2017 08:50:00	130.00	8296.00	15.20	90.63	1.19	8228.00

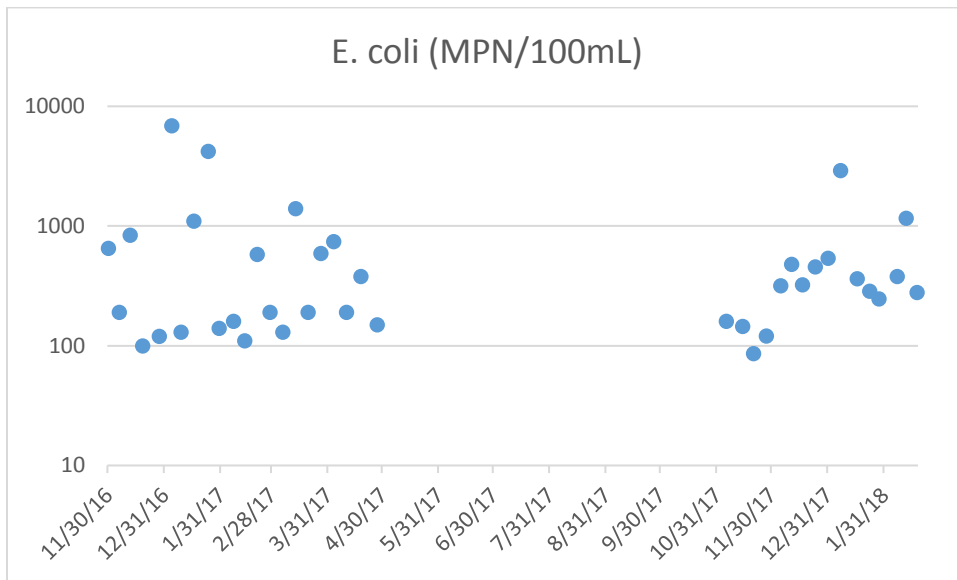
Let's take a look at the E. Coli data. Convert the date-time format to something that is easier to label on charts



Use **Line Chart** and adjust the display to show just points (since these are instantaneous data).



Adjust the vertical axis so that it is a log scale with no decimal places. You can see that the E Coli data given here are sampled intensively during the summer when the swimming season is occurring. They are sampled monthly for the rest of the year but the data are not shown here. You can see that the E Coli value rarely falls below 100 MPN/100mL.



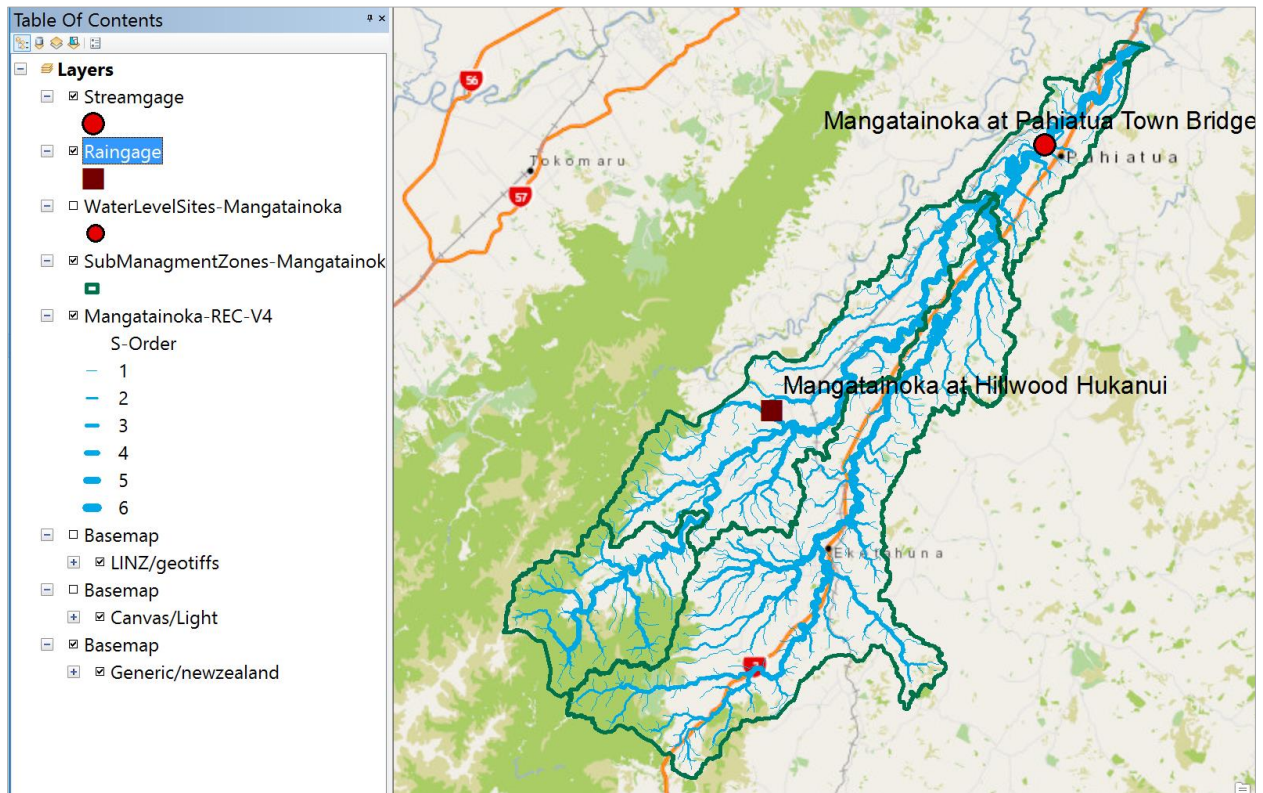
To be Turned In: A plot of E Coli as a function of time for the given data. Determine the mean, median, maximum and minimum value of these data.

Daily Time Series Data

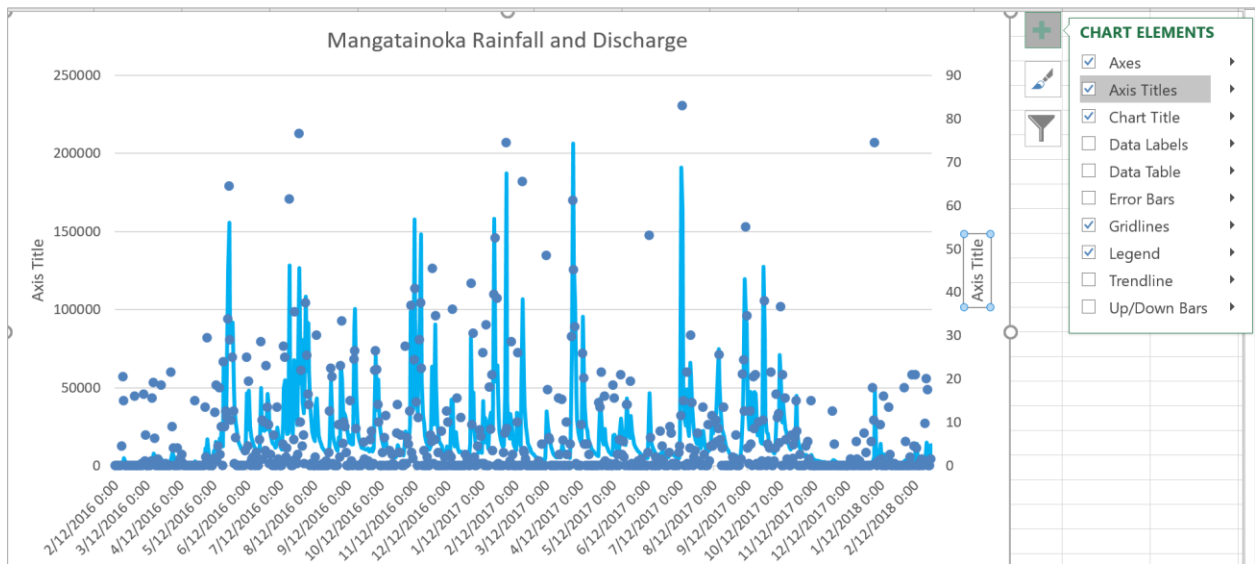
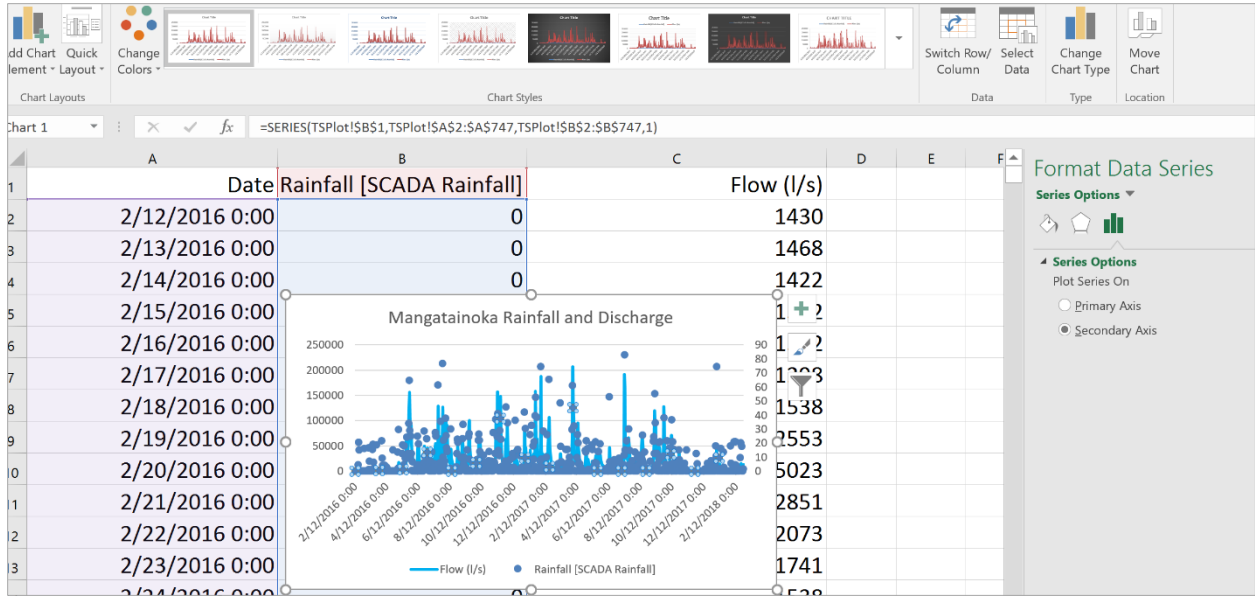
The second spreadsheet [Daily Timeseries](#) contains the daily cumulative rainfall (mm) for the rain gage in the Mangatainoka catchment at Hillwood Hukanui, and the statistical average across

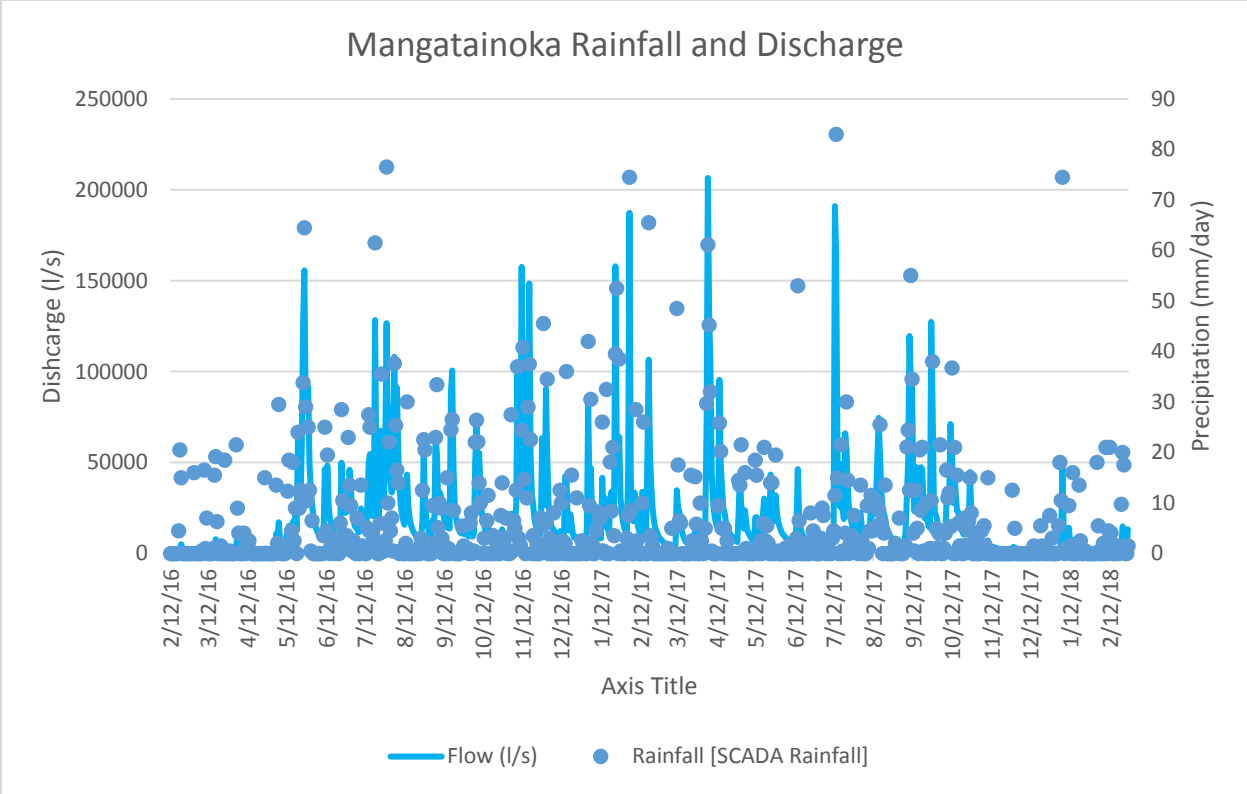
each day of values recorded each XX minutes of flow, water temperature, dissolved oxygen saturation and turbidity. The raingage is near the center of the Mangatainoka catchment, and the other variables are recorded on the Mangatainoka River at Pahiatua Town Bridge. You can see that there are some days where Turbidity readings are incomplete.

Date	Mangatainoka at Hillwor Rainfall [SCADA Rainfall] (mm)	Mangatainoka at Pahiatua To Flow (l/s)	Mangatainoka at Pahiatua Town B Water Temperature (°C)	Mangatainoka at Pahiatua To DO Saturation (%)	Mangatainoka at Pahiatua To Turbidity (ISO) (FNU)
2/12/2016 0:00	0	0	1430	20.8	101.92
2/13/2016 0:00	0	0	1468	20.6	102.61
2/14/2016 0:00	0	0	1422	21.5	105.41
2/15/2016 0:00	0	0	1362	20.9	106.86
2/16/2016 0:00	0	0	1312	20.4	92.29
2/17/2016 0:00	0	0	1298	19.4	95.31
2/18/2016 0:00	4.5	1538	1538	19	93.93
2/19/2016 0:00	20.5	20.5	2553	19.1	92.21
2/20/2016 0:00	15	15	5023	19.2	91.62
2/21/2016 0:00	0	0	2851	19.7	95.35
					20.19
					4.72
					34.35
					68.39
					279.34
					149.8

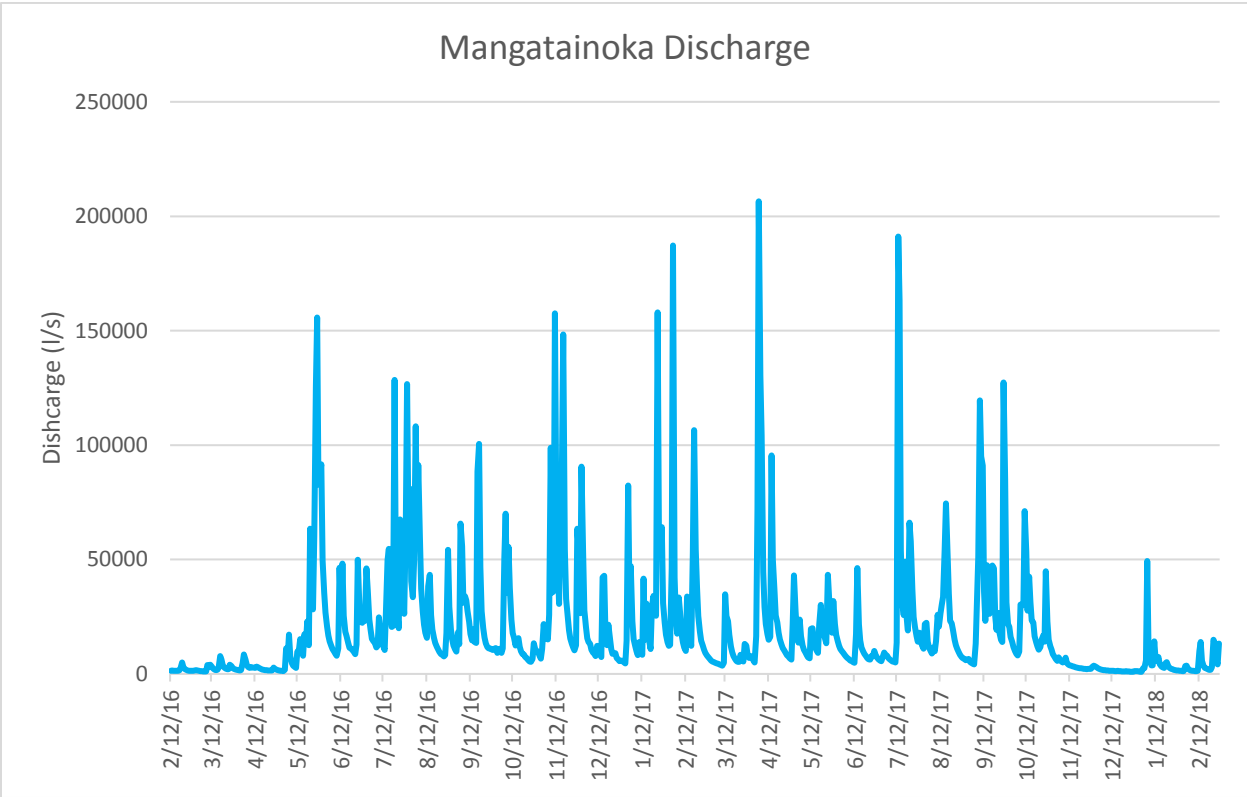


Let's take a look at the daily rainfall and streamflow data.





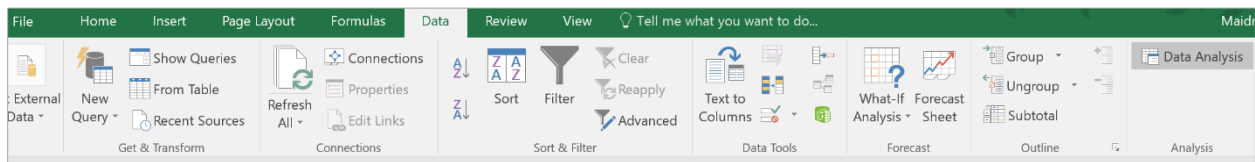
Here is a plot of the Mangatainoka discharge by itself. This charge has a very characteristic pattern. Of rapid rise as rainfall occurs and slower recession after the rainfall stops.



To be Turned In: A plot of rainfall and discharge for a shorter interval so you can see the effect of individual rain events on the discharge at Town Bridge. Plot the Turbidity and Discharge as a function of time so you can see how sediment transport responds to changes in flow. Plot the Water Temperature as a function of time – what is the range of water temperature from summer to winter at this location? Plot the DO saturation as a function of time – what do you notice about the values in the summer?

Regression Modeling in Excel

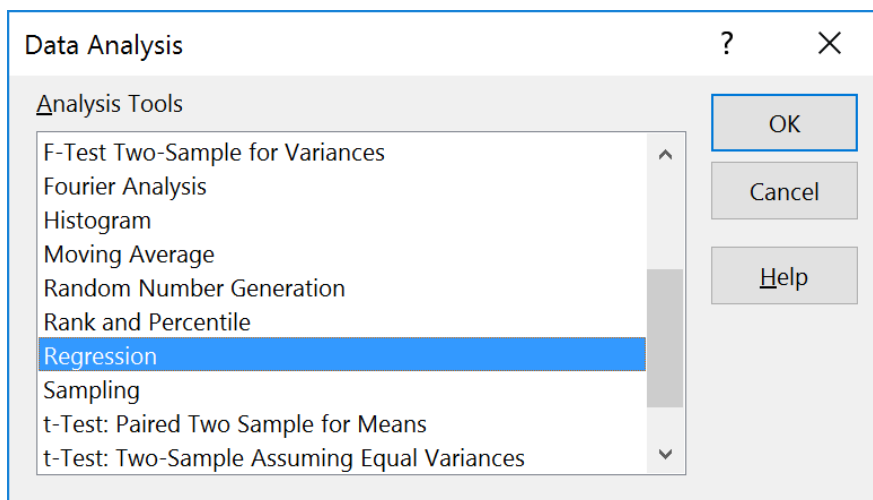
If you click on the **Data** tab in Excel, there is a **Data Analysis Toolpak** that shows up in the right hand side of the display.



If you don't see this toolpak, it means that it has not been loaded for the first time into your version of Excel. Click the File tab, click Options, and then click the **Add-Ins** category. In the Manage box, select **Excel Add-ins** and then click Go. If you're using **Excel** for Mac, in the file menu go to **Tools > Excel Add-ins**. In the **Add-Ins** box, check the **Analysis ToolPak** check box, and then click OK.

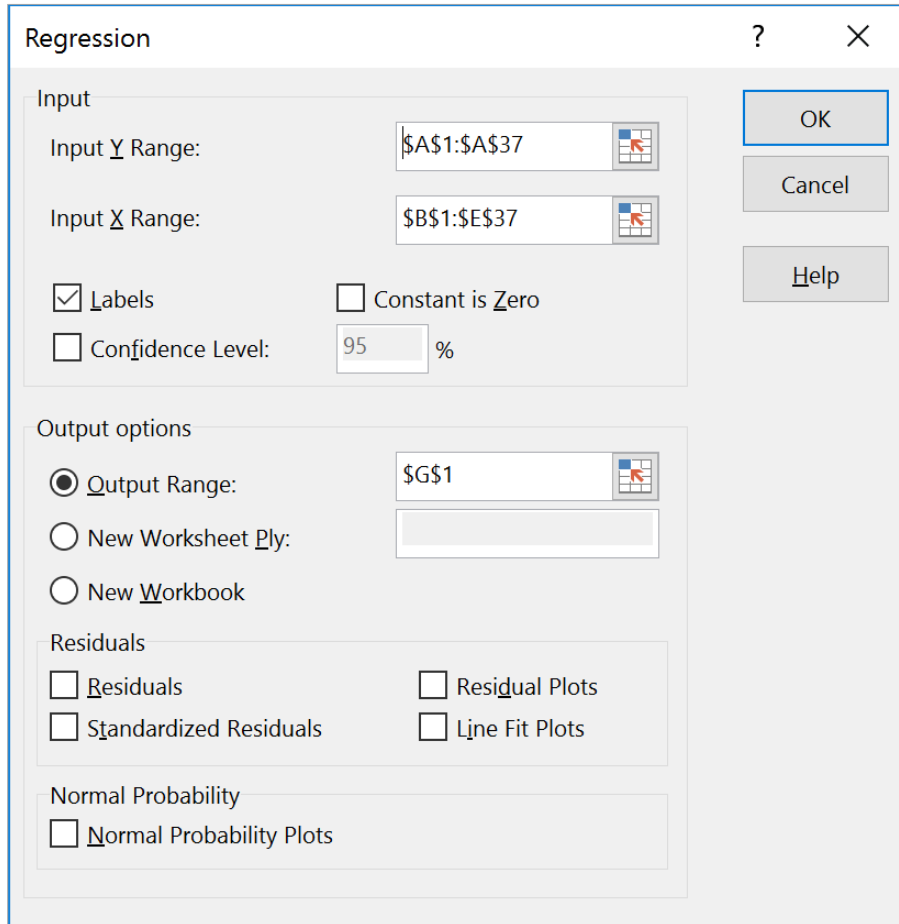
Delete any rows of data for which there are missing values for Turbidity.

Once you have the Analysis ToolPak loaded, click on it and selection **Regression**



Choose the **Input Y range** to cover the values for E Coli, and the **Input X range** to cover the values for the dependent variables (Discharge, Temperature, DO Saturation, and Turbidity). Make sure you check the box marked **Labels** so that you know that the topmost row has your

column labels. Put the top left cell of the **Output range** to be out in the clear space on your Excel page. When you hit Ok, you should see a set of Regression results appear.



Below is the result for a model that includes all four observed variables. The R Square indicates that 64% of the variation in E Coli is accounted for by the model, which is quite reasonable given the large statistical variation in coliform measurements. However the t Stat of the TEMP and DOSAT variables is less than an absolute value of 2, which makes it unlikely these variables contribute significantly to the solution.

SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.802375
R Square	0.643806
Adjusted R Square	0.597846
Standard Error	847.1182
Observations	36

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	4	40208573	10052143	14.00782	1.24E-06
Residual	31	22245888	717609.3		
Total	35	62454461			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	-861.29	3162.618	-0.27233	0.787171	-7311.49	5588.911
DISCHARGE	0.055295	0.011806	4.683562	5.31E-05	0.031216	0.079374
TEMP	69.53101	82.95406	0.838187	0.408342	-99.6549	238.7169
DOSAT	-6.59418	27.303	-0.24152	0.810744	-62.279	49.09064
TURBIDITY	64.46987	18.27121	3.528494	0.001328	27.20549	101.7343

Here is the result for a model that just includes Discharge and Turbidity as explanatory variables. In this instance the R Square is about 63%, which shows that leaving out the Temperature and DO Saturation has made little difference to the result. The t-statistics for Discharge and Turbidity are above an absolute value of 2 which suggests that both variables are statistically important. The predictive two-variable model is then

$$\text{ECOLI} = -256.685 + 0.049438 * \text{DISCHARGE} + 67.90047 * \text{TURBIDITY}$$

(1)

<i>Regression Statistics</i>	
Multiple R	0.795254
R Square	0.632429
Adjusted R Square	
Square	0.610152
Standard Error	834.0562
Observations	36

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2	39498022	19749011	28.38931	6.73E-08
Residual	33	22956439	695649.7		
Total	35	62454461			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	-256.685	190.4553	-1.34774	0.186921	-644.169	130.7995
DISCHARGE	0.049438	0.010075	4.906907	2.43E-05	0.02894	0.069937
TURBIDITY	67.90047	17.53049	3.873277	0.000481	32.23441	103.5665

Here is the result with just discharge, and you can see that the R Square has dropped to 46.5% so this is a less effective model. However, we'll need to use it because on some days the turbidity is not measured and a prediction of EColi using Discharge alone will be needed. In that case, the predictive model will be:

$$\text{ECOLI} = -34.3929 + 0.061779 * \text{DISCHARGE} \quad (2)$$

SUMMARY OUTPUT

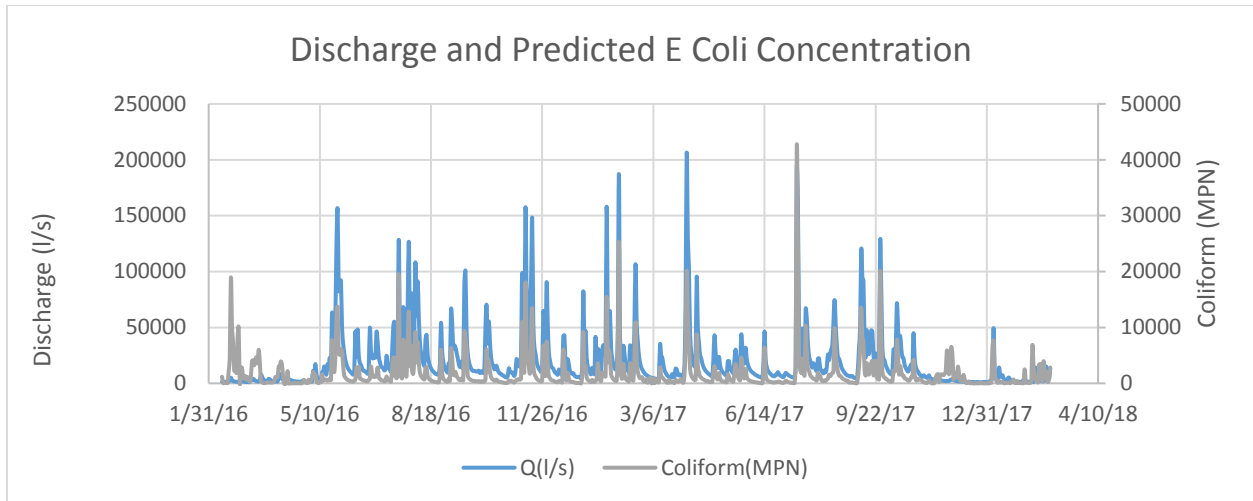
<i>Regression Statistics</i>	
Multiple R	0.682148
R Square	0.465326
Adjusted R Square	0.4496
Standard Error	991.0299
Observations	36

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	29061692	29061692	29.59017	4.61E-06
Residual	34	33392768	982140.2		
Total	35	62454461			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	-34.3929	215.7811	-0.15939	0.874306	-472.913	404.127
DISCHARGE	0.061779	0.011357	5.439684	4.61E-06	0.038699	0.08486

Taking Equations (1) and (2) and applying them to the daily time series data for the Mangatainoka river at Town Lake Bridge for February 2017 to February 2018. In making this computation, I found that on a few occasions the predicted values are negative, so I set those values to 0. The resulting average value over the data period for E Coli concentration is 1812 MPN or 1812 Coliform Units per 100mL of water. This is well above the acceptable limits for contact recreation, so you probably should not swim there. Here is a chart of discharge and predicted Coliform concentrations.



Coliform Loading

The Coliform Load, L , is computed as the product of the discharge, Q , and concentration, C :

$$L = Q * C$$

If you take the observed daily mean discharge (liters/second) and predicted E Coliform concentration (CFU/100mL), I got the resulting product to give an average value of $L = 1.10 * 10^8$ (l/s – CFU/100mL). To resolve the units, 1 liter = $10 * 100\text{mL}$, and 1 day = 86,400 s, so the load, $L = 1.10 * 10^8$ (l/s – CFU/100mL) * 10 (100mL/1Liter) * 86,400 (s/day) = $9.48 * 10^{13}$ CFU/day.

Ok, now you can get the dimensions of the bacterial pollution problem. This means that on average a loading of $\sim 10 \times 10^{12}$ or 10 Billion Billion Coliform Units per day are flowing past the Pahiatua Town Bridge on the Mangatainoka River.

To be turned in: Use Regression to prepare a statistical model for the prediction of E. Coli levels in the Mangatainoka River at Pahiatua Town Bridge. Prepare an estimate of the mean concentration of E. Coli for the year at this location using this model. How does this compare with the mean value calculated from the instantaneous sampled values of E. Coli? On a daily basis, multiply the discharge and estimated concentration to get an E Coli load value and find mean daily load in CFU/day for this location.

To be Turned In

- (1) *A plot of E Coli as a function of time for the given data. Determine the mean, median, maximum and minimum value of these data.*
- (2) *A plot of rainfall and discharge for a shorter interval so you can see the effect of individual rain events on the discharge at Town Bridge. Plot the Turbidity and Discharge as a function of time so you can see how sediment transport responds to changes in flow. Plot the Water Temperature as a function of time – what is the range of water temperature from summer to winter at this location? Plot the DO saturation as a function of time – what do you notice about the values in the summer?*

(3) Use Regression to prepare a statistical model for the prediction of E. Coli levels in the Mangatainoka River at Pahiatua Town Bridge. Prepare an estimate of the mean concentration of E. Coli for the year at this location using this model. How does this compare with the mean value calculated from the instantaneous sampled values of E. Coli in (1)? On a daily basis, multiply the discharge and estimated concentration to get an E Coli load value and find mean daily load in CFU/day for this location.