Virtual Beach 3.0.6 – Data Preparation for MLR model

In this module you will learn how to:

- A. Import and clean-up model-building data for your beach
- B. Process directional data (wind, currents, waves)
- C. Combine two or more predictive variables
- D. Transform variables and explore potential relationships

A. Import and clean-up model-building data for your beach

A.1. Open **Microsoft Excel** to preview the data you will be importing into **Virtual Beach 3**. Open the file "VB_Training_Data_MLR.xls".

Be sure to save your data as "*.xls" files. A plugin for **Virtual Beach** 3 is available for importing "*.xlsx" files, but there are still bugs to be worked out.

Column **B** is always the *response* variable, "ECOLI" in this example. All data to the right are potential *explanatory* variable. See the **KEY** tab of the **Excel** file for descriptions of variables used in this module. Close the **Excel** file before returning to **Virtual Beach 3**. Data cannot be imported from an open Excel file.

l	ਜ਼ •5• ∂ - ∓		1-VB-Tra	aining-Data.xls	[Compatibility Mod	e] - Excel	SONYA C	ARLSON 🖻	- 0	×
F	ile Home	Insert	Page Layout	Formulas	Data Revie	w View 🖓	? Tell me what ye	ou want to do	<u>A</u>	+ Share
B1	• •	׼	fx	ECOLI						~
	А	В	С	D	E	F	G	н	1	
1	DATETIME	ECOLI	QTRSEASON	PRE_JUNE21	JUNE21_JULY15	JULY16_AUG10	POST_AUG10	WATERTEMP_F	DOY	RRA
2	5/21/2009 12:05	2	1	1	0	0	0	62	141	
3	5/28/2009 12:20	5	1	1	0	0	0	55	148	
4	6/4/2009 11:55	1	1	1	0	0	0	65	155	
5	6/11/2009 12:35	345	1	1	0	0	0	56	162	
6	6/12/2009 14:15	18	1	1	0	0	0	66	163	
7	6/15/2009 11:25	29	1	1	0	0	0	62	166	
8	6/16/2009 10:30	8	1	1	0	0	0	68	167	
9	6/17/2009 14:05	120	1	1	0	0	0	64	168	
10	6/18/2009 14:05	17	1	1	0	0	0	68	169	
11	6/22/2009 10:40	4	2	0	1	0	0	70	173	
12	6/23/2009 11:45	76	2	0	1	0	0	75	174	
13	6/24/2009 1.55	15	2	0	1	0	0	72	175	-
	KEY	10	del-Building	data_2009-15	Validation dat	a (+) 🕴 🗄	•			Þ

• This file can be used as a template for formatting beach-specific data.

A.2. Return to Virtual Beach 3 project file created in the "Beach Orientation" module.
1. Click the Global Datasheet tab. 2. Click the Import Data icon and select the Excel file "VB_Training_Data_MLR.xls". 3. Click Open.

Virtua	al Beach 3	Global Datashee	1	Popen				×
•	0		S @		is PC → OS (C:) → Virtual Beach 3	~ ♡	Search Virtual Beach	
Import Data	Validate Data	Compute Manipula A O	te Transform Go To Model	Organize 👻 New folde	er			• 🔳 🔞
Add	12	Work	with Data	Program Files (^	Name	Date modified	Туре	Size
		2		Recovery Temp Users Virtual Beach 3	配 VB_Training_Data_MLR.xls	8/15/2017 6:08 PM	Microsoft Excel 97	667 KB
				File n	ame: VB_Training_Data_MLR.xls	~	Spreads et Files (*.) Open	xls;*.xlsx;*.c ∨ Cancel

A.3. In this example there is more than one worksheet in the Excel file, so you must choose which one to import. **1.** Select the worksheet **MLR-Model-Building data_2009-15. 2.** Click **OK**.

🛃 Virtual	I Beach 3			
File	Location	Global Datasheet		
Import Data Add	Validate Data Validate	Compute A O Work wit	Go To Model	Select Excel Workbook — X Worksheets in VB_Training_Data_MLR.xds KEY
				MLR-Validation data 2009-15 MLR-Validation data 2016 1 0k Cancel

A.4. The data table will open in Virtual Beach 3.



F Virtual Beach 3 automatically highlights the second column of the datasheet as the response variable, "ECOLI", in this example. The "Response Variable" is indicated in the left-hand panel, along with "Column Count", "Row Count" and other descriptions of the data.

Locatio	Glob	al Datasheet									
ort Validate Data dd Validate	AO	C Manipulate Work with		Ga Ta Model							
File		1-VB-Training-	Data :	DATETIME	ECOLI	QTRSEASON	PRE_JUNE21	JUNE21_JULY15	JULY16_AUG10	POST_AUG10	16
Column Count		124		5/21/2009 12:05.	2	1	1	0	0	0	
Now Count		287		5/28/2009 12:20.	5	1	1	0	0	0	
	DATETIME	-	6/4/2009 11:55:	1	1	1	0	0	0		
tesponse vana	bie	ELULI		6/11/2009 12:35.	. 345	1	1	0	0	0	
isabled Row C	ount	0		6/12/2009 2:15:	18	1	1	0	0	0	
Disabled Column		0		6/15/2009 11:25.	. 29	1	1	0	0	0	
Hidden Column Independent Va		0		6/16/2009 10:30.	8	1	1	0	0	0	
nuepenueni, va	indbie Courit	122		6/17/2009 2:05:	. 120	1	1	0	0	0	
				6/18/2009 2:05:	17	1	1	0	0	0	-
				6/22/2009 10:40.	. 4	2	0	1	0	0	
				6/23/2009 11:45.		2	0	1	0	0	-
	_	_	> 0	6/24/2009 11:55	15	2	n	1	n	n	1×

Virtual Beach 3 will NOT build a model if any cells have null (missing), or non-A.5. numeric (text) values. 1. Click the Validate Data icon to check your dataset. 2. In the pop-up window, click Scan.

Location ort Validate Data	Data Validation							
d Validate	(Optional) Find:	ECOLI	QTRSEASON	PRE_JUNE21	JUNE21_JULY15	JULY16_AUG10	POST_AUG10	_
nie Column Count	(e) Replace With:	2	1	1	0	0	0	T
olumn Count	O Delete Row	5	1	1	0	0	0	-
ate-Time Index	O Delete Column	0	1		-	1 A	-	-
Response Variable		1	1	1	0	0	0	
	Take Action Within	345	1	1	0	0	0	
isabled Row Count	Only This Cell \sim	18	1	1	0	0	0	
isabled Column Count		29	1	1	0	0	0	
fidden Column Count ndependent Variable Co	Tenz Adding	8	1	1	0	0	0	
idependent valiable oc		120	1	1	0	0	0	
		17	Ì	1	0	0	0	
	Identify Categorical Variables	4	2	0	1	0	0	
		70	2	n	1	n	n	

A.6. In this example, the "CLOUDCOV_qual" column is flagged because the values are text, or non-numeric. **1.** Click the radio button next to **Delete Column**. Under **Take Action Within** make sure **Only This Column** is selected. **2.** Click **Take Action**.

	Validation						- 0	1
Location	Data Validation Scan							
ta Data Id Validate	(Optional) Find:							
File		QTRSEASON	PRE_JUNE21	JUNE21_JULY15	JULY16_AUG10	POST_AUG1	CLOUDCOV_Qual	
Column Count	Replace With:	1	1	0	0	0	Sunny	1.
Row Count	O Delete Row	1	1	0	0	0	Cloudy	1
ate-Time Index	Delete Column	1	1	0	0	0		
Response Variable	Take Action Witmin:	1	1	0	0	0	Mostly Sunny	1
Disabled Row Count	Only This Column 🗸 🗸	1	1	0	0	0	Cloudy	1
isabled Column Count		1	1	0	0	0	Sunny	1
dden Column Count dependent Variable C	Take Action	1	1	0	0	0	Cloudy	1
		1	1	0	0	0	Sunny	1
		1	1	0	0	0	Sunny	1
			0	1	0	0	Sunny	1
	Identify Categorical Variables	- Z						

A.7. Repeat step A.6 until you come to the "TRIB6" column. The variable is numeric, but some cells are empty. Do not remove the entire column. **1.** Click the radio button for **Delete Row**. **2.** Select **Entire Column** and click **Take Action**.

Virtual Beach 3	Validation						-	
Location	Data Validation							
ort Validate Cor ta Data	Scan							
d Validate	(Optional) Find:							
File	Action:	RRAIN24	RRAIN48	RRAIN72	RRAIN120	RRAIN144	TRIB6	T
Column Count	O Replace With:	0	0	0	60.27	63.52	813	92
Row Count.	Delete Row	0.88	0.88	0.88	12.76	62.9	571.7	60
Date-Time Index	O Delete Column	0	0.88	0.88	0.88	12.76	586.5	57
Response Variable	Take Action Within:	0	0	22.51	22.51	23.39	333.1	40
Disabled Row Count	Only This Row	0	0	0	22.51	22.51	193.3	23
Disabled Column Count	Only This Row	0.38	0.38	0.38	22.89	22.89	200.8	19
Hidden Column Count Independent Variable Co	Entire Column Entire Sheet	0	0.38	0.38	0.38	22.89	206.2	20
ndependent vanable et	hr	1.63	12.52	12.52	12.52	12.9	169.6	17
		0	1.63	12.52	12.52	12.52	161.4	16
	Identify Categorical Variables	2 16	16	17.63	28.52	28.52		
<			10	10	20 62	20 23	1	
	Reform Cancel							
Location Global Da								

A.8. Repeat Step A.7 until a notice appears at the bottom of the pop-up window stating **No anomalous data values found**. Then click the **Return** button.

ort Validate Co a Data	Validation Data Validation						
d Validate	Scan						
File	(Optional) Find:	ECOLI	QTRSEASON	PRE_JUNE21	JUNE21_JULY15	JULY16_AUG10	POST_AUG10
Column Count	Action	345	1	1	0	0	0
Row Count	Replace With:	. 18	1	1	0	0	0
Date-Time Index Response Variable	(b) Delete Row	29	1	1	0	0	0
hesponse variable	Delete Column	. 17	1	1	0	0	0
Disabled Row Count		4	2	0	1	0	0
Disabled Column Count	Take Action Within	76	2	0	1	0	0
Hidden Column Count Independent Variable (Entire Column	15	2	0	1	0	0
	Take Actor	3	2	0	1	0	0
	Tank na u	11	2	0	1	0	0
	^	. 5	2	0	1	0	0
	Identify Categorical Variables	. 7	2	0	1	0	0
	2	26	2	n	1	n	0
	No anomalous data values found	1					

B. Process wind and current data

B.1. **1.** Click the **Compute A O** icon. **2.** In the pop-up window, under **Wind Data**, click the pull down arrow next to **Speed** and select WSPD. For **Direction**, select WDIR.

Virtual Beach 3	Wind/Current Co	omponents — 🗆 🗄	×					
Location Global	Wind Data Specify wind data	columns:						
ort Validate Compute ta Data A O Id Valida		WVPD24 A WVPAR WVPERP		_				
File T		WVPAR3 WVPERP3 WVPAR6	QTRSEASON	PRE_JUNE21	JUNE21_JULY15	JULY16_AUG10	POST_AUG10	D ^
Column Count S	Current Data	WVPERP6	1	1	0	0	0	14
Row Count 2	Specify current dat	WVPAR12	1	1	0	0	0	14
Date-Time Index	Specily culteric da	WVPAR24	1	1	0	0	0	15
Response Variable E	Speed	WVPERP24 WSPD	1	1	0	0	0	16
Disabled Row Count		WDIR	1	1	0	0	0	16
Disabled Column Count	Direction (deg)	WPAR WPERP	1	1	0	0	0	16
Hidden Column Count		WPAR3 WPERP3	1	1	0	0	0	16
ndependent Variable Count		WPAR6	-		0	0	0	16
the second se		WPERP6 WPAR12	-	1		-	-	-
	Specify wave data	WPERP12	1	1	0	0	0	16
		WPAR24	2	0	1	0	0	17
	Height	WPERP24 ATEMP ATEMP6	2	n	1	n	n	17 4
ocation Global Datashe	Direction (deg)	ATEMP12 ATEMP24 ATEMPstdv6	-					

B.2. Repeat for **Current Data**, selecting CSPD and CDIR. Repeat for **Wave Data** selecting WVHT and WVDIR. The Beach Angle is automatically included. Click **OK**.

Current Data 3 0.1261 -0.01659 0.1717 -0.03284 0.1625 -0.00388 Specify current data columns: -0.0162 0.0718 -0.2873 0.09819 -0.437 0.1281 Specify current data columns: -0.0162 0.03684 -0.1361 0.03951 -0.1784 0.04436 Direction (deg) CDIR -0.0051 0.04907 -0.1737 0.04087 -0.1929 0.02522 13 -0.131 0.04138 -0.146 0.04777 -0.1977 0.06138 09 0.06425 -0.009418 0.09981 0.004245 0.1182 -0.01211 187 -0.1005 -0.01455 -0.08743 -0.01437 -0.1853 0.03237 99 -0.209 0.05336 -0.1797 0.04137 -0.1853 0.03237 187 -0.1005 -0.0133 -0.04925 -0.01564 -0.05114 0.004655 187 -0.02085 0.01331 -0.04001 0.02431 -0.01950 0.00567 11 <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th>Wind Data</th>									Wind Data
Direction (deg) WDIR P3 CPAR6 CPERP6 CPAR12 CPERP12 CPAR24 CPERP24 3 0.1261 -0.01659 0.1717 -0.03284 0.1625 -0.03281 Current Data								columns:	Specify wind data
P3 CPAR6 CPERF6 CPAR12 CPERP12 CPAR24 CPERP24 Current Data 3 0.1261 -0.01659 0.1717 -0.03284 0.1625 -0.03688 Specify current data columns: 8 -0.1912 0.0718 -0.2873 0.09819 -0.437 0.1281 Specify current data columns: -0.102 0.03684 -0.1361 0.03951 -0.1784 0.04436 Direction (deg) CDIR -0 0.06425 -0.09418 0.04907 -0.1289 0.02522 13 -0.1331 0.04138 -0.146 0.04777 -0.1977 0.06125 -0.01211 187 -0.1055 -0.0145 -0.08743 -0.004245 -0.01349 -0.01155 09 -0.2094 0.05336 -0.1797 0.04137 -0.1833 0.03237 Wave Data Specify wave data columns: - - - - 0.00536 -0.1797 0.04137 -0.1833 0.03237 F22 -0.004242 -0.01033	 						-	WSPD ~	Speed
Current Data 8 0.1912 0.0718 0.2873 0.09819 0.437 0.1281 Specify current data columna: 5 0.102 0.03684 0.1361 0.03951 0.1784 0.0436 Direction (deg) CDIR 2 0.04087 0.1281 0.04087 0.1281 0.04087 0.1281 0.04436 Wave Data CDIR 2 0.06425 0.009418 0.09981 0.04087 0.1182 0.01151 Wave Data Specify wave data columns: 2 0.0005 0.01331 0.04001 0.02431 0.01839 0.02536 Wave Data Specify wave data columns: 2 0.0004242 0.01331 0.04001 0.02431 0.07802 0.02366 127 0.0004242 0.01331 0.04001 0.02431 0.01055 0.00697 118 0.0004213 0.01828 0.03929 0.004252 0.04437 0.02096 126 0.02442 0.01331 0.04001 0.02431 0.01055 0.00697 <t< td=""><td>CPERP24</td><td>CPAR24</td><td>CPERP12</td><td>CPAR12</td><td>CPERP6</td><td>CPAR6</td><td>RP3</td><td>WDIR ~</td><td>Direction (deg)</td></t<>	CPERP24	CPAR24	CPERP12	CPAR12	CPERP6	CPAR6	RP3	WDIR ~	Direction (deg)
Specify current data columna: 5 0.102 0.03684 0.1361 0.03951 0.1784 0.0436 Specify current data columna: 0 0.2051 0.04907 0.1737 0.04087 0.1289 0.02522 Direction (deg) CDIR 0 0.06425 0.00918 0.09981 0.04077 0.1977 0.01182 0.01182 Wave Data 0 0.00536 0.01045 0.08743 0.004235 0.01211 187 0.1005 0.01045 0.08743 0.004774 0.1823 0.03237 9 0.209 0.05336 0.1797 0.04137 0.1853 0.03237 57 0.02085 0.01031 0.04001 0.02431 0.07802 0.02666 722 0.0004242 0.01033 0.04925 0.01564 0.05114 0.004655 266 0.02442 0.007306 0.04628 0.004371 0.01095 0.00697 01 0.07245 0.00895 0.1101 0.00833 0.1157 0.009645	-0.03638	0.1625	-0.03284	0.1717	-0.01659	0.1261	3		
Specify current data columns. 0.2051 0.04907 0.1737 0.04087 0.1289 0.02522 Direction (deg) CDIR 0 0.0311 0.04138 0.146 0.04777 0.1977 0.06138 Wave Data 0 0.06425 0.009418 0.09811 0.004245 0.1182 -0.01211 187 0.1005 -0.01045 -0.08743 -0.004734 -0.01349 -0.0155 09 -0.209 0.05336 -0.1797 0.04137 -0.1853 0.03237 57 -0.02085 0.00131 -0.04001 0.02431 -0.07802 0.02666 722 -0.004242 -0.01033 -0.04925 -0.01564 -0.05114 0.004655 226 0.02442 0.007306 0.04628 0.004371 0.01095 0.00697 01 0.07245 -0.00895 0.1101 -0.008333 0.1157 -0.02096 48 -0.0024213 0.01828 0.03929 0.004252 0.04437 0.008455	 0.1281	-0.437	0.09819	-0.2873	0.0718	-0.1912	8		Current Data
Speed CSPD -0.2051 0.04907 -0.1737 0.04087 -0.1289 0.02522 13 -0.1331 0.04138 -0.146 0.04777 -0.1977 0.06138 Direction (deg) CDIR - - - - - - - - - - - - - - - - 0.04087 - - - - - - - 0.04087 - 0.1289 0.02522 - - - - - - - 0.1977 0.01977 0.01977 0.01977 0.01974 - 0.01134 - 0.01245 0.01134 - 0.01473 - 0.0183 0.03237 57 - 0.0205 0.01331 - 0.04001 0.02431 - 0.02366 722 - 0.0004242 - 0.01331 - 0.04001 0.02431 - 0.02455 - 0.01544 - 0.02146 0.004252	0.04436	-0.1784	0.03951	-0.1361	0.03684	-0.102	5	ta columos:	Specify current da
Direction (deg) CDIR 13 0.1331 0.04138 0.146 0.04777 0.09777 0.06138 Direction (deg) CDIR 09 0.06425 -0.009418 0.09981 0.004245 0.1182 -0.01211 187 -0.1005 -0.01045 -0.08743 -0.004734 -0.01349 -0.01155 09 -0.209 0.05336 -0.1797 0.04137 -0.1853 0.03237 Wave Data 57 -0.02085 0.001331 -0.04001 0.02431 -0.07802 0.02366 Specify wave data columns: - - - - - - 0.004242 - - 0.04332 - 0.01351 - 0.04371 0.01095 0.02665 126 0.02442 -0.01033 -0.04925 -0.01564 -0.05114 0.00657 01 0.07245 -0.008955 0.1101 -0.008333 0.1167 -0.02096 48 -0.0004213 0.01828 0.03929 0.004252 0.04437	 0.02522	-0.1289	0.04087	-0.1737	0.04907	-0.2051			
Wave Data -0.0005 -0.01045 -0.08743 -0.004734 -0.01349 -0.01155 Specify wave data columns: -0.000420 0.0209 0.05336 -0.1797 0.04137 -0.1853 0.03237 Height WVHT -0.0004242 -0.01033 -0.04925 -0.01564 -0.05114 0.00465 10 0.07245 -0.00895 0.1101 -0.008393 0.1167 -0.02096 48 -0.0004213 0.01828 0.03239 0.04437 -0.04437 -0.008355	0.06138	-0.1977	0.04777	-0.146	0.04138	-0.1331	13	CSPD ~	Speed
Wave Data -0.005 -0.0145 -0.08743 -0.01479 -0.01349 -0.01155 Dy ave Data 0.0209 0.05336 -0.1797 0.04137 -0.1853 0.03237 Specify wave data columns:	-0.01211	0.1182	0.004245	0.09981	-0.009418	0.06425	09	1	
Wave Data 57 -0.02085 0.001311 -0.04001 0.02431 -0.07802 0.02366 Specify wave data columns: Height -0.0004242 -0.01033 -0.04925 -0.01564 -0.05114 0.00452 Leight WVHT -0.07245 0.00895 0.1101 -0.00833 0.1167 -0.00964 Direction (dep) WVDIR -0.0004213 0.01828 0.03929 0.004252 0.04437 -0.008545	-0.01155	-0.01349	-0.004734	-0.08743	-0.01045	-0.1005	187		Direction (deg)
Specify wave data columns: 722 -0.0004242 -0.01033 -0.04925 -0.01564 -0.05114 0.00465 Height WVHT 0 0.02442 0.007306 0.04628 0.004371 0.01095 0.00697 Direction (dep) WVDIB 0 0.002421 0.01828 0.03929 0.04252 0.04437 -0.009645	0.03237	-0.1853	0.04137	-0.1797	0.05336	-0.209	09		
Understand User	0.02366	-0.07802	0.02431	-0.04001	0.001331	-0.02085	57		Wave Data
Height WVHT 226 0.02442 0.007306 0.04628 0.004371 0.01095 0.00697 Direction (dea) WVDIB 0.07245 -0.00895 0.1101 -0.008333 0.1167 -0.02966 48 -0.0004213 0.01828 0.03929 0.004252 0.04437 -0.008545	0.00465	-0.05114	-0.01564	-0.04925	-0.01033	-0.0004242	722	a columns:	Specify wave data
Direction (dep) WV/DIR 01 0.07245 -0.00855 0.1101 -0.00833 0.1167 -0.0296 WV/DIR 48 -0.0004213 0.01828 0.03929 0.004252 0.04437 -0.008545	0.00697	0.01095	0.004371	0.04628	0.007306	0.02442	226		
Direction (deg) WVDIR	-0.02096	0.1167	-0.008393	0.1101	-0.00895	0.07245	01	WVHT ~	Height
Direction (deg) WVDIR	-0.008545	0.04437	0.004252	0.03929	0.01828	-0.0004213	48	hanna	-
07 -0.0003726 0.05574 -0.03808 0.04123 -0.02744 0.04941	0.04941	-0.02744	0.04123	-0.03808	0.05574	-0.0003726	07	MADIK ~	Direction (deg)
9 -0.2809 0.1204 -0.2872 0.1212 -0.1852 0.1123	0.1123	-0.1852	0.1212	-0.2872	0.1204	-0.2809	9		
Reach Angle (deg): [21.1542739868164 02 -0.2167 0.07601 -0.1695 0.0632 -0.222 0.0754	0.0754	-0.222	0.0632	-0.1695	0.07601	-0.2167	02	Angle (deg): 21.1542739868164	Beach

B.3. Scroll to the far-right end of the table. <u>Six</u> new columns have been added to the end of the global data sheet and that the unprocessed wind, current, and wave data columns are now inactive (red text):

Wind A_comp: along-shore wind speed Wind O_comp: toward shore wind speed Current A_comp: along-shore current speed Current O_comp: toward shore current speed Wave_A_comp: along-shore wave height Wave_A_comp: on-shore wave height

ort Validate Data	Compute A O	Manipulate Work with	Transform	Go To Model								
File		1-VB-Training-I	Data:	CPI	ERP24	WindA_comp[WDII	WindO_comp[WDI	CurrentA_comp[CD	CurrentO_comp[CD	WaveA_comp[WVI	WaveO_comp[WV	-
Column Count		126		1.0	2638	5.224	1.006	0.09794	0.01167	0.2734	0.2603	
Row Count		281		0.12	81	-0.2000		0.1070	0.0774		-0.1104	1
Date-Time Index		DATETIME		0.04	436	-0.8499	-2.613	-0.07477	-0.03144	-0.07845	-0.06703	1
Response Variabl	le	ECOLI		0.02	522	-8.08	-3.52	-0.2327	-0.06925	-0.6112	0.01645	1
Disabled Row Co	unt	0		0.06	138	-4.677	-1.274	-0.1248	-0.04007	-0.3093	0.0552	1
Disabled Column		6		-0.0	1211	-1.941	0.4918	0.0493	0.02722	-0.01497	0.01758	1
Hidden Column C ndependent Varia		0 118		-0.0	1155	0.04079	-2.718	-0.0708	-0.01735	-0.1998	-0.002227	
nuependent van	able count	110		0.03	237	-1.62	-2.762	-0.2096	-0.04655	-0.3067	0.06824	1
1		_	> <	H		+ + × r					· · · · · · · · · · · · · · · · · · ·	-

C. Combine two or more predictive variables

Interaction Terms: In situations where two predictive variables are themselves correlated, meaning they interact with one another in terms of how they influence water quality, it may be beneficial to combine them into a single interaction term by **multiplying** them together. Combined the two variables may be better predictors of water quality than if included individually.

Combined Categories: Some variables are either yes or no. The 0 is "no" and the 1 is "yes". In situations where binary variables represent successive categories of some qualitative variable, like visually-observed water clarity, it may be useful to combine them into a single binary variable by **summing** them. The resulting variable will have a value of 1 when *either* of the two conditions is present. This can be especially helpful when there is little functional distinction between the categories or few cases in which one of the conditions is ever observed. In this example, the difference between TURBID and OPAQUE water is not very distinct; if the water is turbid, it was probably also opaque.

Change-in-Flow Variables. In situations where continuous stream flow data are preprocessed over different timeframes, **subtracting** one temporal snapshot from another can create proxy variables for *changes* in flow. The difference between 24-hour maximum and minimum flow rates indicates whether recent tributary discharge has been consistent or very different after a flash flood event.

C.1. First create an interaction term by multiplying two variables together. 1. Click the Manipulate icon. 2. In the pop-up window, ctrl-select WAVEHEIGHT_FT and GULLS.
3. Click the right-arrow ">" button.

	20 14	Build Expression						
		Independent Variable	s	Variables in	Expression			
port Validate (ata Data dd Validate	Manipulate T Work with I	QTRSEASON PRE_JUNE21 JUNE21_JULY15 JULY16_AUG10 POST_AUG10	^ >		3			
File	1-VB-Training-Di	WAVEHEIGHT FT			0.00		0.11	0
Column Count	126	AIRTEMP_F		• Sum	O Diff O N	hax O Min	() Mean	O Product
Row Count	281	GULLS		-				
Date-Time Index	DATETIME	SUNNY	Z	-				
Response Variable	ECOLI	MSUNNY PSUNNY MCLOUDY		Add	Remove		2nd On	der Interaction
Disabled Row Count	0	CLOUDY						
Disabled Column Cou	nt 6	CLARITY						
Hidden Column Count	0	CLEARWATER						
	Count 118	TURBID	~					

C.2. 1. Click the radio button next to **Product**. 2. Click the **Add** button.

ta Data A O Validate Work with I PRE_JUNE21 JULY15 JULY16_AUG10 POST_AUG10 POST_AUG10 WATERTEMP_F AIRTEMP_F AIRTEMP_F Column Count 126 Row Count 281 Date-Time Index DATETIME PSUNNY MCLOUDY CLOUDY CLARITY Claarity CLOUDY Claarity CLOUDY Disabled Column Count 6		*	24	Build Expression			
All Dock PRE JUNE21 Validate Work with I PRE JUNE21 JULY15 JULY15 JULY15 JULY15 JULY15 JULY16 AUG10 POST_AUG10 POST_AUG10 POST_AUG10 WATERTEMP_F Row Count 126 SUNNY SUNNY Date-Time Index DATETIME PSUNNY MCLOUDY CLOUDY CLARWATER Disabled Row Count 0 Disabled Column Count 6 Hidden Column Count 0 OPAQUE OPAQUE			e Manipulate T		1. X	_	
File 1-VB-Training-D: AIRTEMP F Column Count 126 Sunny Sunny Row Count 281 Date-Time Index DATETIME PSUNNY Response Vanable ECOLI Disabled Row Count 0 Disabled Column Count 6 STURBID Hidden Column Count 0		AO	Work with 0	PRE_JUNE21 JUNE21_JUL	Y15		SULLS
File 1-VB-Training-D. AIRTEMP F Column Count 126 CLOUDCOV Row Count 281 MSUNNY Date-Time Index DATETIME PSUNNY Response Variable ECOLI MCLOUDY CLARITY Disabled Row Count 0 CLEARWATER SURNER Disabled Column Count 6 STURBID TURBID Hidden Column Count 0 OPAQUE				POST_AUG1	٥	<	
Column Count 126 SUUNYY Row Count 281 MSUNNY Date-Time Index DATETIME PSUNNY Response Variable ECOLI MCLOUDY CLARITY CLARITY Add Disabled Row Count 0 CLEARWATER Disabled Column Count 6 STURBID Hidden Column Count 0 OPAQUE	File		1-VB-Training-Di	AIRTEMP_F	-		Sum O Diff O Max O Min O Mann @ Double
Date-Time Index DATETIME PSUNNY Response Variable ECOLI MCLOUDY CLOUDY CLARITY Add Disabled Row Count 0 CLEARWATER Disabled Column Count 6 STURBID TURBID Hidden Column Count 0	Column Count		- 6 T D				
Response Variable ECOLI MCLOUDY CLOUDY CLARITY Add Remove 2nd Order Interaction Disabled Row Count 0 CLEARWATER 2 Disabled Column Count 6 STURBID 2 Hidden Column Count 0 OPAQUE 2							PROD[WAVEHEIGHT_FT,GULLS]
Response Vanable ELULI CLOUDY CLARITY Add Remove 2nd Order Interaction Disabled Row Count 0 CLEARWATER 2 2 2 Disabled Column Count 6 STURBID 2 2 Hidden Column Count 0 OPAQUE 2							
Disabled Row Count 0 CLEARWATER Disabled Column Count 6 STURBID Hidden Column Count 0 OPAQUE	Response Variabl	e	ECOLI	CLOUDY			Add Remove 2nd Order Interaction
Hidden Column Count 0 OPAQUE			0	CLEARWATE	R		
Hidden Column Count 0 OPAQUE	Disabled Column	Count	6				Z
						_	
	Independent Varia	able Count	118		ORE	~	

C.3. This creates an interaction term that may describe more accurately how wave height influences the number of gulls on the beach. **1.** Shift-select WAVEHEIGHT_FT and GULLS. **2.** Click the left-arrow "(" button to move them back to the main list.

-	Global Datasheet				
ort Validate Data d Validate	Compute A O Work with	PRE_JUNE21 JUNE21_JULY15	^	WAVEHEIGHT_FT GULLS	
File 1-VB-Training- Column Count 126 Row Count 281 Date-Time Index DATETIME		JULY16_AUG10 POST_AUG10 WATERTEMP_F AIRTEMP_F CLOUDCOV SUNNY MSUNNY		Sum O Diff O Max O Min	O Mean Product
Date-Time Index Response Variable	DATETIME ECOLI	PSUNNY MCLOUDY CLOUDY	2		2nd Order Interaction
Disabled Row Coun Disabled Column Co Hidden Column Cou Independent Variabl	nunt 6 Int 0	CLARITY CLEARWATER STURBID TURBID OPAQUE ALGNEARSHORE	~	PROD[WAVEHEIGHT_FT,GULLS]	

C.4. Repeat the steps in C.2 and C.3 to create the expression of combined categories TURBID and OPAQUE using the **Sum** radio-button. This combined expression creates a variable where a visual observation of either TURBID or OPAQUE water receives a value of 1.

ild Expression				
ndependent Variables		Variables in Expression		
GULLS CLOUDCOV SUNNY MSUNNY	> <	TURBID OPAQUE		
PSUNNY MCLOUDY CLOUDY CLARITY CLEARWATER		Sum O Diff O Max O Min	O Mean O Product	
STURBID		SUM[TURBID,OPAQUE]		
ALGNEARSHORE ALGNR_NONE ALGNR_LOW		Add Remove	2nd Order Interactions	
ALGNR_MOD ALGNR_HIGH ALGBEACH ALGBCH_NONE ALGBCH_LOW ALGBCH_MOD		PROD[WAVEHEIGHT_FT,GULLS] SUM[TURBID.OPAQUE]		

C.5. Repeat the steps in C.2 and C.3 to create the change in flow variable with TRIBmax24 and TRIBmin24 using the **Diff** radio-button. This approximates whether and to what extent the previous 24 hours of tributary discharge has been constant or varied a lot. Other manipulations can be added as needed. Click "OK" when complete.

Independent Variables TRIB24 TRIB48 TRIBmax48 TRIBmin48	^		
TRIB72 TRIB168 TRIB336 LAKELEV LAKELEV12 LAKELEV24 WVHT3		○ Sum	
WVHTmax3 WVHT6 WVHTmax6 WVHT12 WVHT12 WVHTmax12 WVHTmax24 WVHTmax24 WVPD	~	Add Remove 2nd Order Interactions PROD[WAVEHEIGHT_FT,GULLS] SUM[TURBID.OPAQUE] DIFF[TRIBmax24,TRIBmin24]	

C.6. Scroll to the far-right end of the table to see any new columns added through this process.

Location Globa	Datasheet										
h Validate Compute a Data A O	Manipulate Tran		Go To Model								
Validate	WORK WITH Dat	a)							
le	1-VB-Training-Data	1	bi c	urrentA complCD	CurrentO_comp[CD	WaveA compIWVI	WayeO comp///	PROD[WAVEHEIG	SUMITURBID OPA	DIFFITRIBmax24 1	Τ,
olumn Count	129		-				0.2603	Less.	0	36	i
	281	1		1678	-0.0771		-0.1104	2700		30	4
	DATETIME		-0	07477	-0.03144	-0.07845	-0.06703	1.2.2.2	0	23	-
esponse Variable	ECOLI			2327	-0.06925		0.01645	45	1	156	-
isabled Row Count	0		-0.	1248	-0.04007	-0.3093	0.0552	125	0	142	1
sabled Column Count	6 0		0.0	0493	0.02722	-0.01497	0.01758	137.5	0	11	-
idden Column Count			-0.	0708	-0.01735	-0.1998	-0.002227	200	0	21	1
Independent Variable Count 121	121		-0.	2096	-0.04655	-0.3067	0.06824	400	1	50	-
			0.0	04153	0.01672	0.07274	0.07508	50	0	37	1
			-0.	004248	-0.02261	-0.09014	0.02469	8	0	20	•

D. Transform variables

D.1. Right-click on the "ECOLI" column header and select Transform > Log10.

To build a usable nowcast model, bacteria counts **<u>must</u>** be transformed. Log10 is a common transformation for microbial concentrations.

ort Validate Comput ta Data A O		sform	Go To Model									
File	1-VB-Training-Data.		DATETIME	ECOLI	Transform	1 3	Log10	Y15	JULY16_AUG10	POST_AUG10	٧A	
Column Count	129		5/21/2009 12:05	2	View Plot		Ln		0	0	62	
Row Count	281		5/28/2009 12:20	5	UnTransf		Power		0	0	5!	
Date-Time Index DATETIME Response Variable ECOLI			6/4/2009 11:55:	1			FOWE		0	0	65	
		6/11/2009 12:35	345	Define Transform >		0		0	0	56		
Disabled Row Count	Count 6 count 0		6/12/2009 2:15:	18	1	1	0		0	0	66	
Disabled Column Count			-	6/15/2009 11:25	29	1	1	0		0	0	6.
			6/16/2009 10:30	8	1	1	0		0	0	68	
			6/17/2009 2:05:	120	1	1	0		0	0	64	
			6/18/2009 2:05:	17	1	1	0	_	0	0	68	
	>	<	6/22/2009 10:40	4	2	0	1		0	0	7(~	

D.2. In addition to transforming the response variable, transforming explanatory variables can significantly improve model fit. **1.** Click the **Transform** icon. **2.** Check all options, EXCEPT Polynomial, and type 0.6666 next to **General Exponent**. **3.** Click **Go**.

As of July 2017, the polynomial transformation causes problems in the Virtual Beach program. Future updates will address this issue.

	al Datasheet		a .		🖶 Transforms to Perform	- 🗆 ×
ort Validate Comput ta Data A O Id Validate	e Manipulate Transt Work with Data	form (Goro Model	_(Available Transforms	Dependent Variable: ECOLI
File	1-VB-Training-Data:		DATETIME	ECO	Square	UG10
Column Count	129		5/21/2009 12:05	2	2⊡ SquareRoot	Go
Row Count	281		5/28/2009 12:20	5		
Date-Time Index	DATETIME		6/4/2009 11:55:	1	QuadRoot	h
Response Variable	ECOLI		6/11/2009 12:35	345	Polynomial	Cancel
Disabled Row Count	0		6/12/2009 2:15:	18	General Exponent 3666	2
Disabled Column Count	6				Select All	3
Hidden Column Count	0		6/15/2009 11:25	29		
Independent Variable Count	121		6/16/2009 10:30	8		
			6/17/2009 2:05:	120		
			6/18/2009 2:05:	17		
			6/22/2009 10:40	4	2 0	1 0

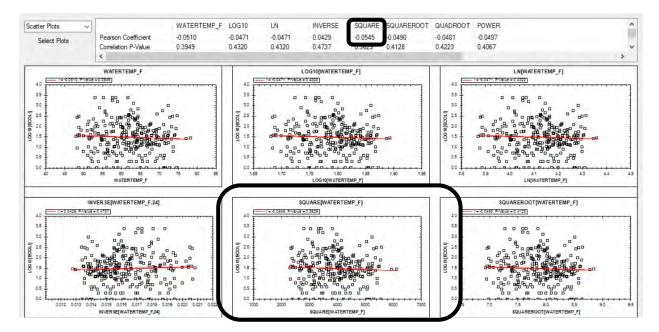
D.3. A pop-up will open listing all of the optional transformations for each explanatory variable. Those in black represent the transformation with the best correlation (Pearson's coefficient) with the response variable LOG(ECOLI).

	Variable	Transform	· · · · · · · · · · · · · · · · · · ·
•	QTRSEASON	none	1
	QTRSEASON	LOG10[QTRSEASON]	1
	QTRSEASON	LN[QTRSEASON]	1
	QTRSEASON	INVERSE[QTRSEASON,0.5]	
	QTRSEASON	SQUARE[QTRSEASON]	6
	QTRSEASON	SQUAREROOT[QTRSEASON]	1.1
	QTRSEASON	QUADROOTIQTRSEASONI	1
	QTRSEASON	POWER[QTRSEASON.0.6666]	
	WATERTEMP_F	none	
	WATERTEMP_F	LOG10[WATERTEMP_F]	
	WATERTEMP_F	LN[WATERTEMP_F]	
	WATERTEMP F	INVERSEIWATERTEMP F.241	1
	WATERTEMP_F	SQUARE[WATERTEMP_F]	
	WATERTEMP_F	SQUAREROUT[WATERTEMP_F]	

D.4. Right-clicking in any cell to the left of a variable will enable you to view scatter plots of all the transformations of that variable versus LOG(ECOLI). For example, right-click to the left of the variable WAVEHEIGHT_FT.

Variables, possible variable interactions, and their	Dependent Varia	ble: LOG10[ECOLI]	
ransforms are shown. Select variables for further processing and modeling.	Variable	Transform	·
	QTRSEASON	none	4
Auto-Select	QTRSEASON	LOG10[QTRSEASON]	() () () () () () () () () ()
The variable or one of its	QTRSEASON	LN[QTRSEASON]	1
transforms is selected by	QTRSEASON	INVERSE[QTRSEASON,0.5]	
maximum Pearson Coefficient. (This is the default view shown.)	QTRSEASON	SQUARE[QTRSEASON]	1
	QTRSEASON	SQUAREROOT[QTRSEASON]	1
Go	QTRSEASON	QUADROOT[QTRSEASON]	1
	QTRSEASON	POWER[QTRSEASON,0.6666]	1
Threshold Select	\sim		
Select a transformed variable only	View Plots	none	
if its Pearson Coefficient exceeds the untransformed variable's	view Plots	LOG10[WATERTEMP_F]	
Pearson Coefficient by a	WATERTEMP_F	LN[WATERTEMP_F]	
specified threshold	WATEDTEND F	INDICOCCOMPTEND COM	

D.5. Note that in this case the best transformation, in terms of Person's r, is square transformation. The scatter plot confirms this selection. Close the window to return to the list of transformation options.



If you decide to select an alternative best transformation for a given variable, simply click on that row. When you are finished, click **OK**.

D.6. New columns are added if the newly-transformed variable had a better fit than untransformed original variable. The new columns here are

SQUARE(WATERTEMP_F), QUADROOT(WAVEHEIGHT_FT), and

QUADROOT(AIRTEMP_F). The original columns are now disabled as indicated by red text. Disabled columns will NOT be used in the model. Save your project file. You can now move onto the next module, "Building an MLR Model".

File	1-VB-Training-Data:		WATERTEMP_F	SQUARE[WATERTEMP_F]	WAVEHEIGHT_FT	QUADROOT[WAVEHEIGHT_FT]	AIRTEMP_F	QUADROOTIAIRT	G			
Column Count	226		62.1	2050	12	1.861	87.1	3.055	50			
Row Count	281					55.4	3069	12	1.861	60.1	2.784	22
Date-Time Index Response Variable	DATETIME LOG10[ECOLI]		65.1	4238	1.5	1.107	62.8	2.815	25			
Response vanable	LOGIN[ECOLI]		56.1	3147	1.5	1.107	61	2.795	30			
Disabled Row Count	Column Count 97 Column Count 1		66.1	4369	0.5	0.8409	57.5	2.754	25			
Disabled Column Count			61.5	3782	0.5	0.8409	65.3	2.843	27			
Independent Variable Count			68.1	4638	1	1	71.7	2.91	20			
			64.3	4134	2	1.189	71.3	2.906	20			
			68.2	4651	1	1	64.3	2.832	50			
< .	>	<	100		1				>			