Exercise I
Excel Part 1 – Power Pivot and Power View

CEE412/CET 522 Transportation Data Management and Visualization

Winter 2020

Introduction

◦ The purpose of this exercise is to get you familiar with data model and visualization functions in Excel, and so we will start with a simple example with two tables of data.

◦ We will create relationships and perform some simple analysis in Power Pivot.

◦ Then using Power View, we will generate a series of summary plots describing high level features of the dataset.

◦ What is a relationship in this context?

◦ A relationship between two tables is generally defined by a set of columns containing values which are present in both of the two tables to be related. Thus, we can find rows in table A that correspond to rows in table B by matching the values of one or more columns (example on the next slide).

Introduction

◦ Suppose I have two tables, one describing employees and one describing their expenses. The relationship between these two tables is based on the employee ID number, named ID in the employee table and Employee_ID in the expenses table. Thus, I can find all expenses for any person by looking up their ID number in the expenses table, and then retrieve their name by looking up the same ID number in the employees table.

Relationship is based on
correspondence between these fields

ID	Name	Title
112	Todd Watson	Professor
232	Judy Jenkins	Research Associate
221	Craig Cook	Student
146	Nicole Flores	Student

Employees

Transaction	Employee_ID	Date	Amount
11007	112	4/8/2015	$22.09
34366	112	2/2/2015	$401.00
34568	146	9/6/2016	$90.99

Expenses

Step 1: Take a Look at the Data

◦ Download the Excel file from Canvas:

E1_Datamodel.xlsx

Take a look at the data:

◦ Note that there are two worksheets, Accidents and Vehicles.

◦ Accidents contains individual accident records for State Route 167, all from 2012.

◦ Vehicles contains a description of the vehicles involved in the accidents in three roles: primary (1), secondary (2), and tertiary (3) (as designated by the VehicleRole column).

◦ Note that the vehicles table has more entries than the accidents table, because more than one vehicle can be involved in each accident.

◦ We can find out which accident each vehicle was involved in by looking up the Report Number from the Accidents table in the Vehicles table.

Step 1: Take a Look at the Data

◦ Most of the fields in the Accidents and Vehicles tables are self explanatory. In particular, we are interested in the following:

  ◦ QUARTER: quarter of the year, that is, quarter 1 is January, February, and March

  ◦ MILE POST WHOLE: binned milepost values, essentially just milepost rounded to the nearest integer

  ◦ VEH COMP DIR TO: An indicator of the travel direction of the primary vehicle in the accident

  ◦ MV DRIVER INJURY TYPE: Type of injury sustained by the driver of a vehicle

  ◦ REPORT NUMBER: A unique identifier for an accident, based on the report which describes the accident

  ◦ MV DRIVER AGE GROUP: binned driver age. For example, 20-29 means the driver was between 20 and 29 years of age inclusive

  ◦ WEATHER: Just that, weather conditions

  ◦ MV DRIVER GENDER: Gender of the vehicle driver

Step 2: Creating Tables

○ Highlight all of the data in the Accidents worksheet and click **Format as Table** in the Home menu.

○ In the resulting dialog box, make sure that the correct data range is selected (see below), and click OK.

○ This will format the active sheet as a table, so that we can reference is as a data source later.
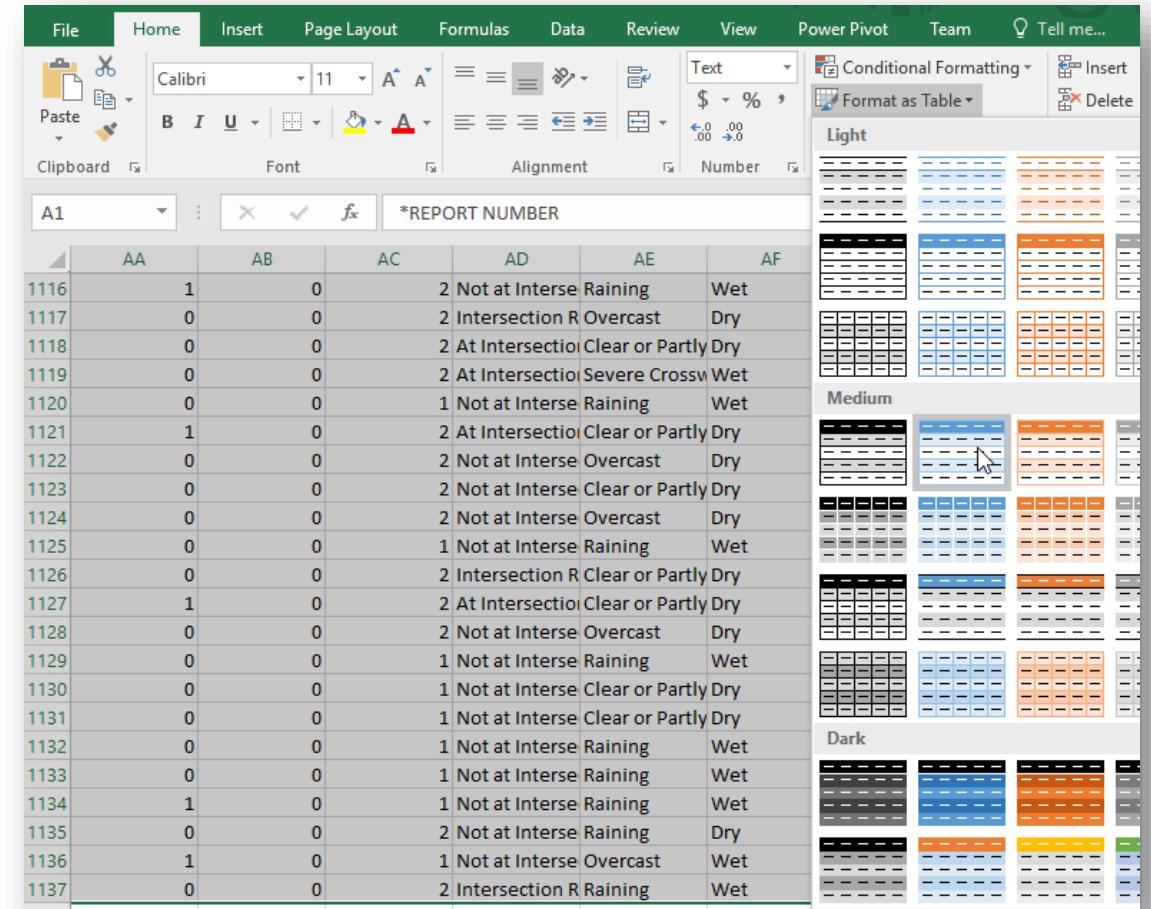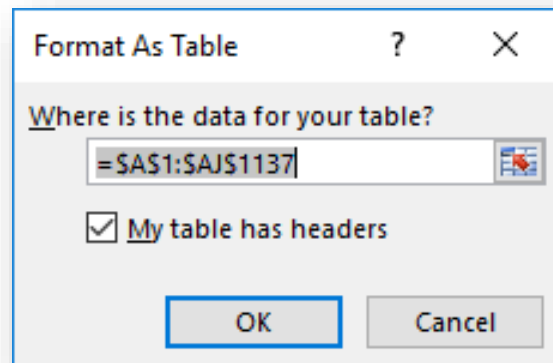
Step 2: Creating Tables

◦ Repeat the process from the previous slide to create a Vehicles table (check the data range on the right).

◦ For each of the two tables you have created, click anywhere inside the table, which will activate the DESIGN tab.

◦ Rename the tables "Accidents" and "Vehicles" in the Design tab by changing the name in the upper left hand corner under Properties (see image on the right).

Step 3: Create Power View Report

- Under the Insert menu, click **Power View** as shown. This will create a Power View report to which we can add the tables we have created.

- If you can't find the Power View button, this is because it has been removed from the ribbon in Excel 2016. Follow steps in the next slide to customize the ribbon and turn on Power View.
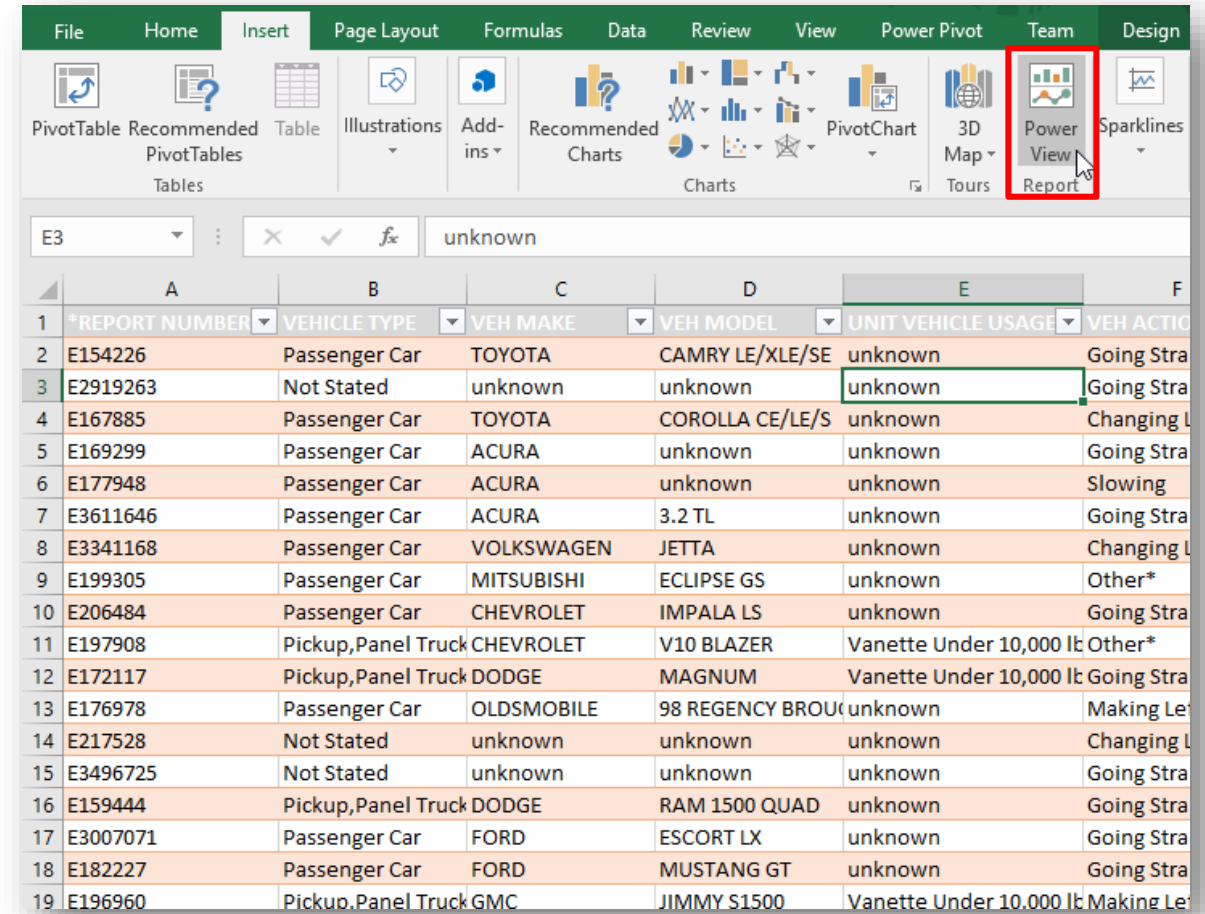
Turn on Power View in Excel 2016 (Optional)

- Click **File** tab → **Options** to open Excel Options window.
- Click **Customize Ribbon**. Under **Main Tabs**, click the plus sign next to **Insert** to expand the Insert tab.
- Click **New Group** to add a new group in the tab.
- A new group will appear in the list. You can give it a name you like (I used "Report").
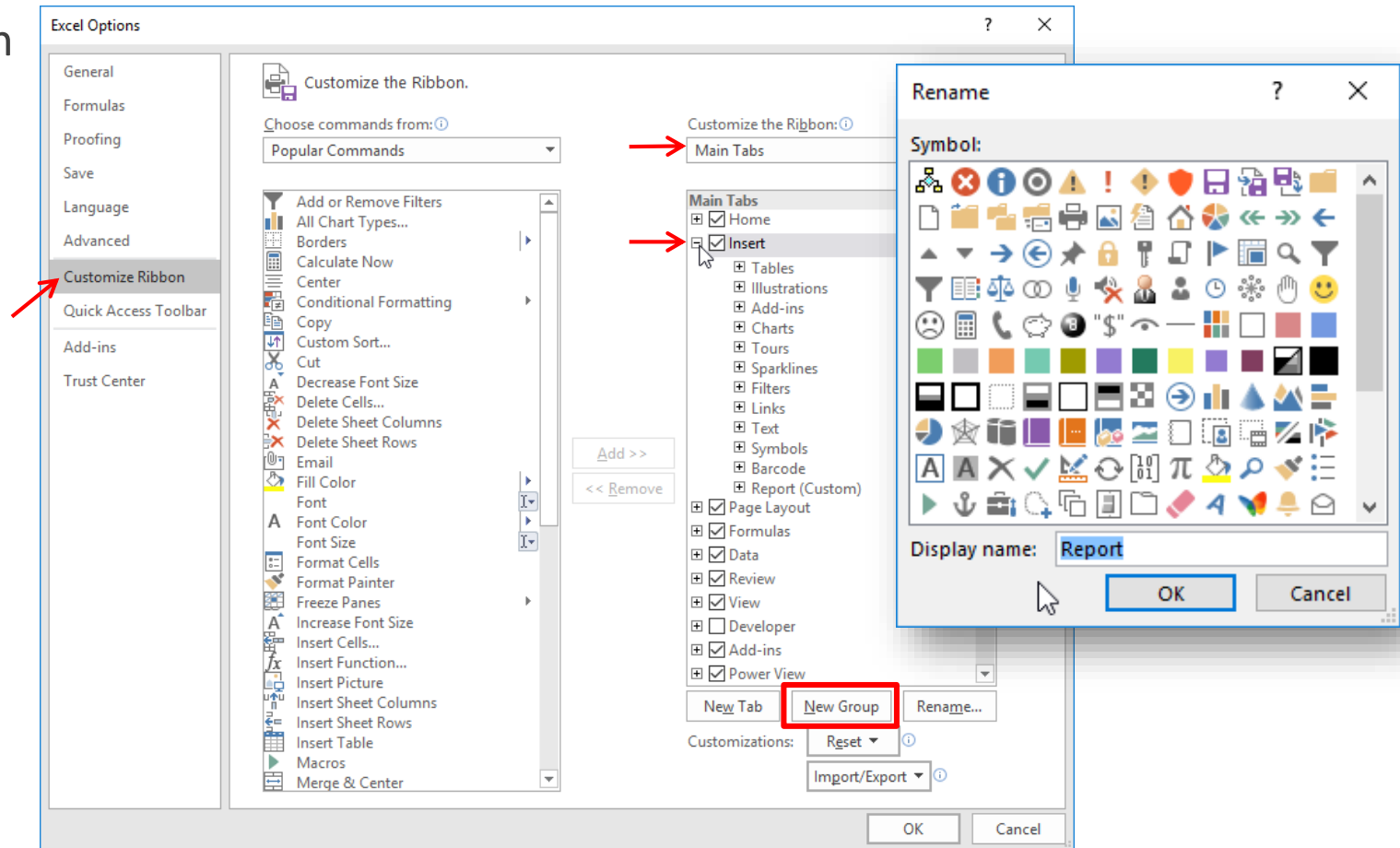
Turn on Power View in Excel 2016 (Optional)

○ Under **Choose commands from**, click **Commands Not in the Ribbon**, and then in the list, pick **Insert a Power View Report**.

○ With both **Insert a Power View Report** and your custom group selected, click **Add**, and then move the your custom group to where you want it on the ribbon.

○ Click **OK** to close the options window.

○ Now you will find a **Power View** button in your **Insert** tab. Click it to create a Power View report.

○ The first time you insert a Power View sheet, Excel prompts you to turn on the Power View add-in. Click **Enable**.
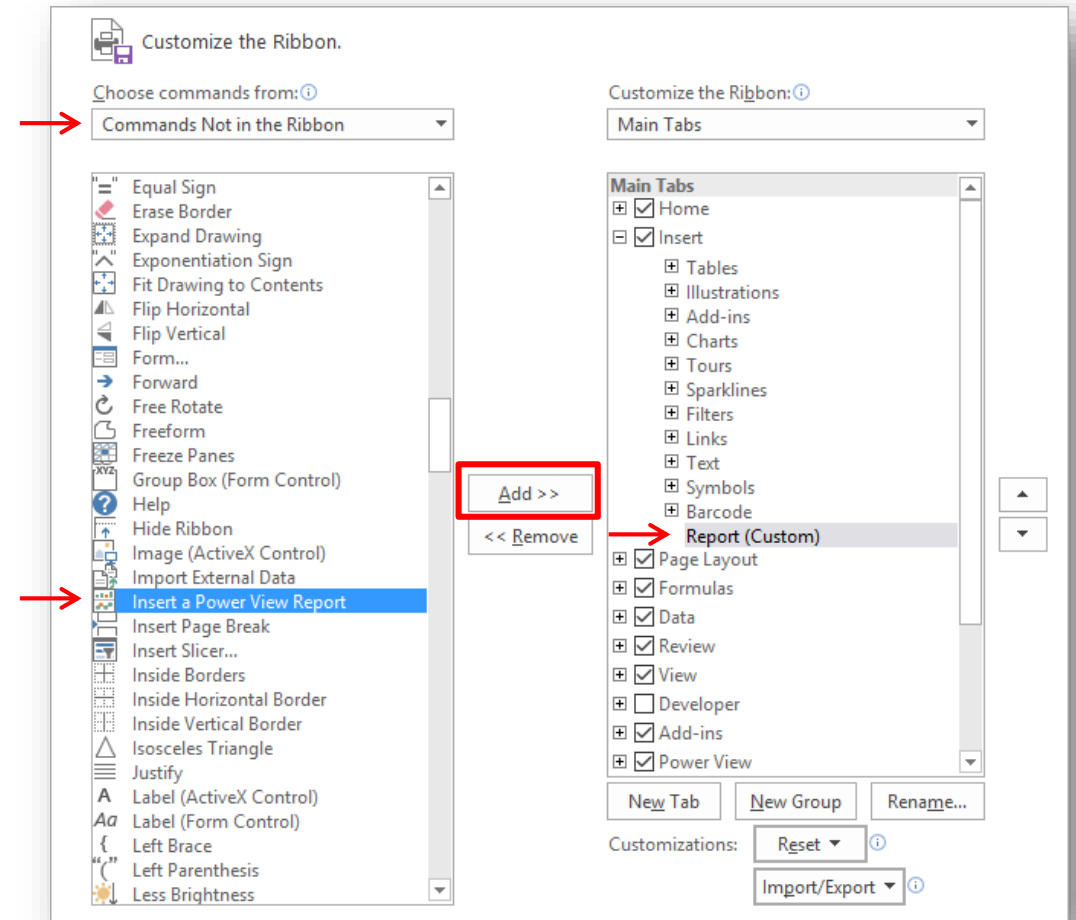
Step 3: Create Power View Report

○ The image on the right shows the new Power View Report that has been created. Note that some fields from your tables have been added by default, remove them by un-checking the boxes near the fields in the Power View Fields menu as shown.

○ Note also that Power View has chosen a data type for each of the fields based on the data that field contains, indicated by the icon next to the associated text box (Σ for numeric, etc.).

Choose which fields to include in the report

Power View Report

Filter area

Step 4: Create Relationships

○ Relationships can be defined in multiple ways. One way is to use the relationship manager window under **Power View** tab.

○ Here we will use a more intuitive way enabled in the PowerPivot diagram view.

○ Note that a **Power Pivot** menu also appears in Excel ribbon after you have enabled the Power View add-on.

○ Click **Manage** to open the Power Pivot window.

○ You will see the data tables in your current data model. This is very similar to a regular Excel window.

○ Under **Home** menu, click **Diagram View**.
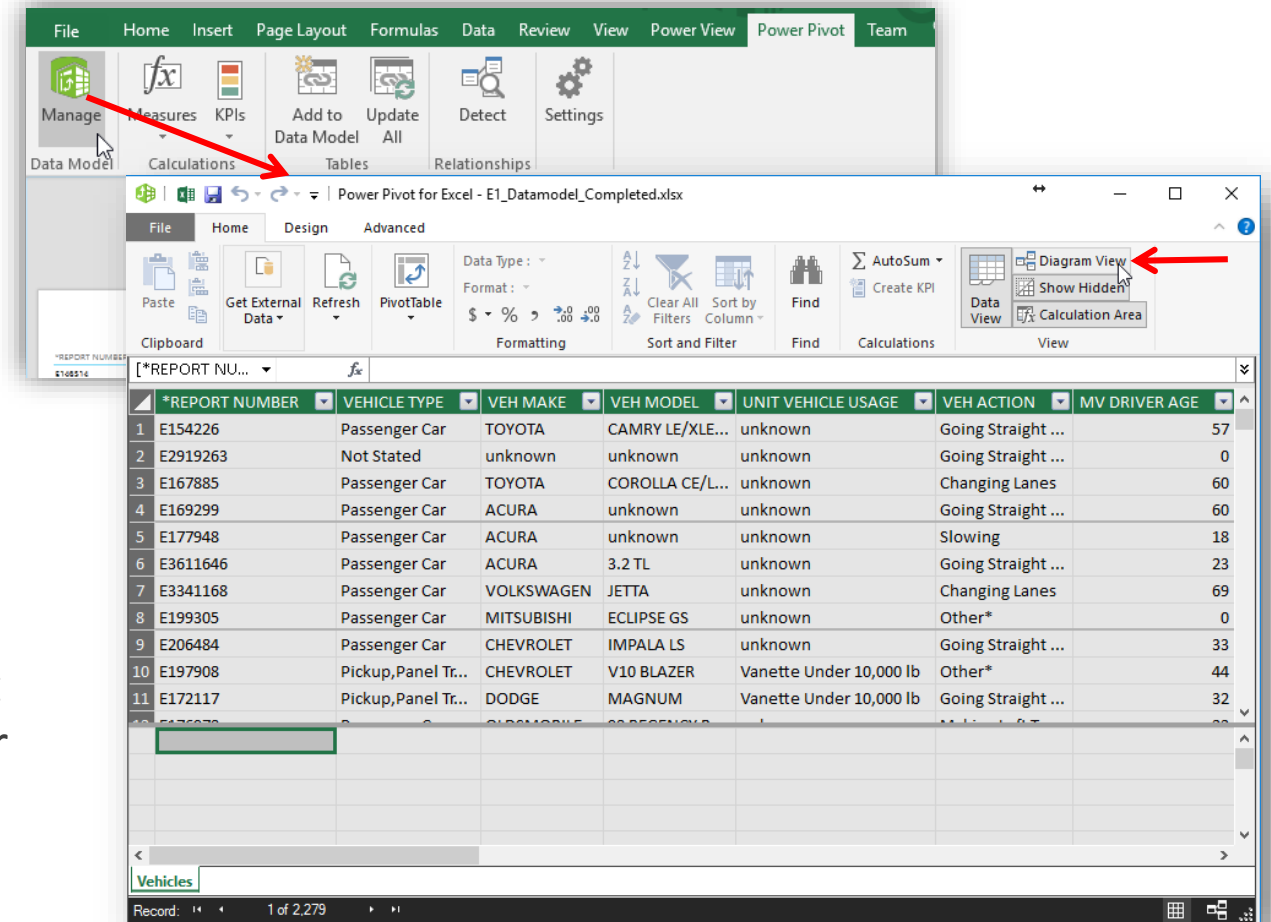
# Step 4: Create Relationships

○ You may only see one table in your data model diagram. This is because the other table has not been connected to the table on which you created the Power View Report.

○ Back to the Excel main window, navigate to the table not included in your data model, click **Add to Data Model** in **Power Pivot** menu.

○ You will see now both tables are in your Power Pivot diagram.
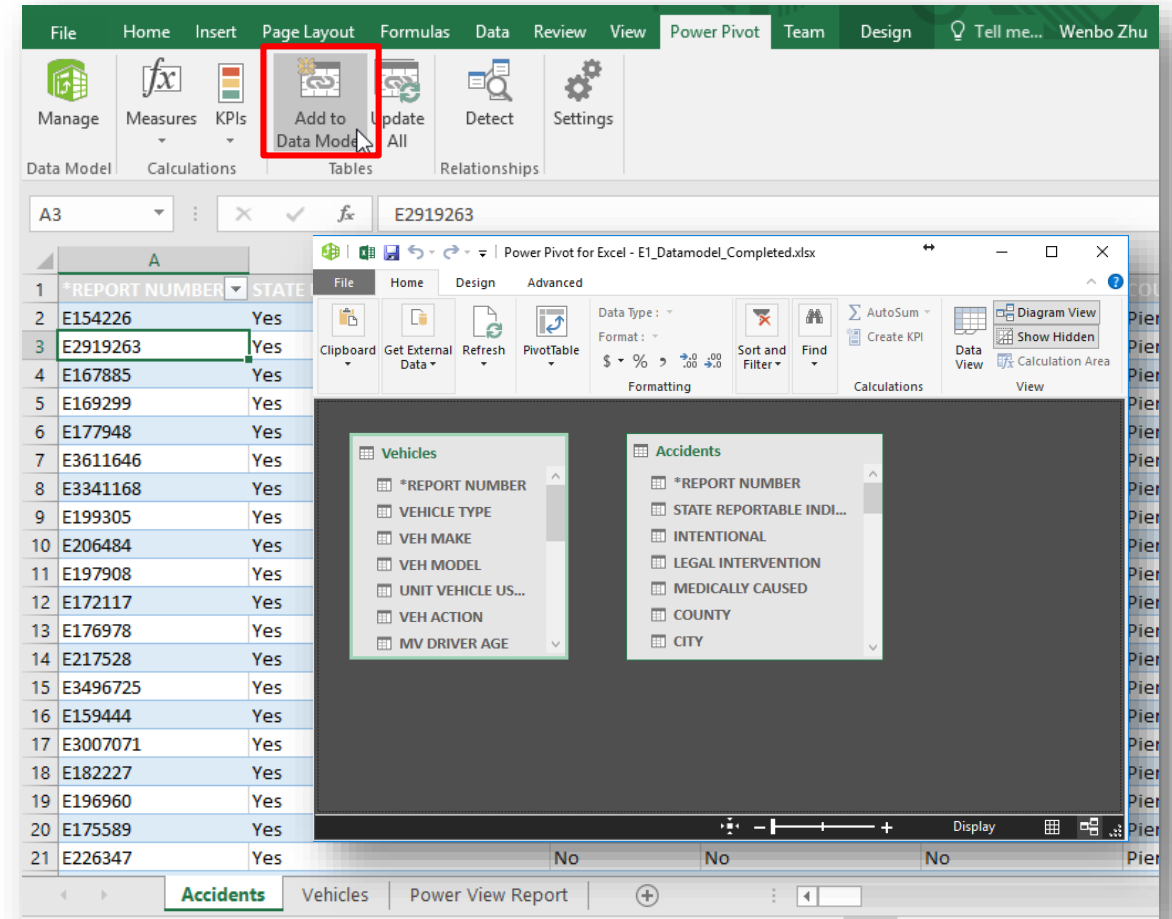
Step 4: Create Relationships

◦ In Power Pivot diagram view, you can create a relationship by simply dragging the join field from one table to the other.

◦ In our data model, the two tables are related by the accident report number.

◦ Find the "*REPORT NUMBER" field in both tables, drag one field to the other (in either direction) as shown.

◦ You will see a line shows up connecting these two tables, with an asterisk (*) on the Vehicles table and a "1" on the Accidents table. This represents Excel auto-detects that this is a many-to-one relationship.
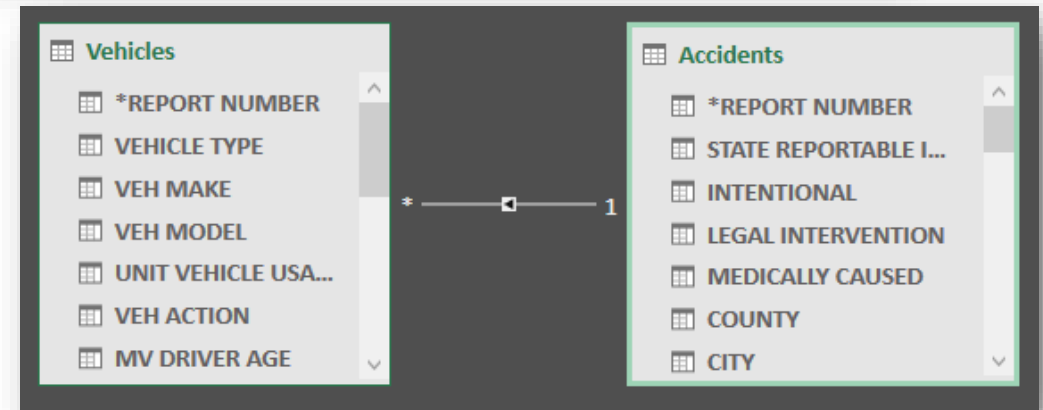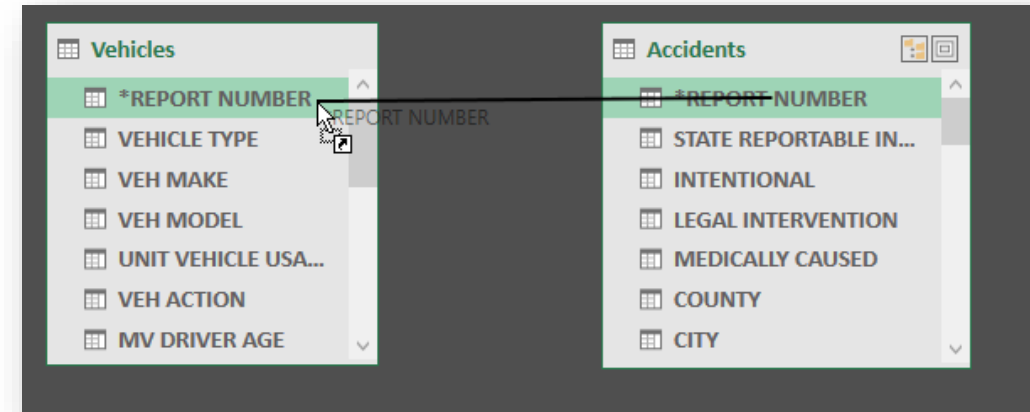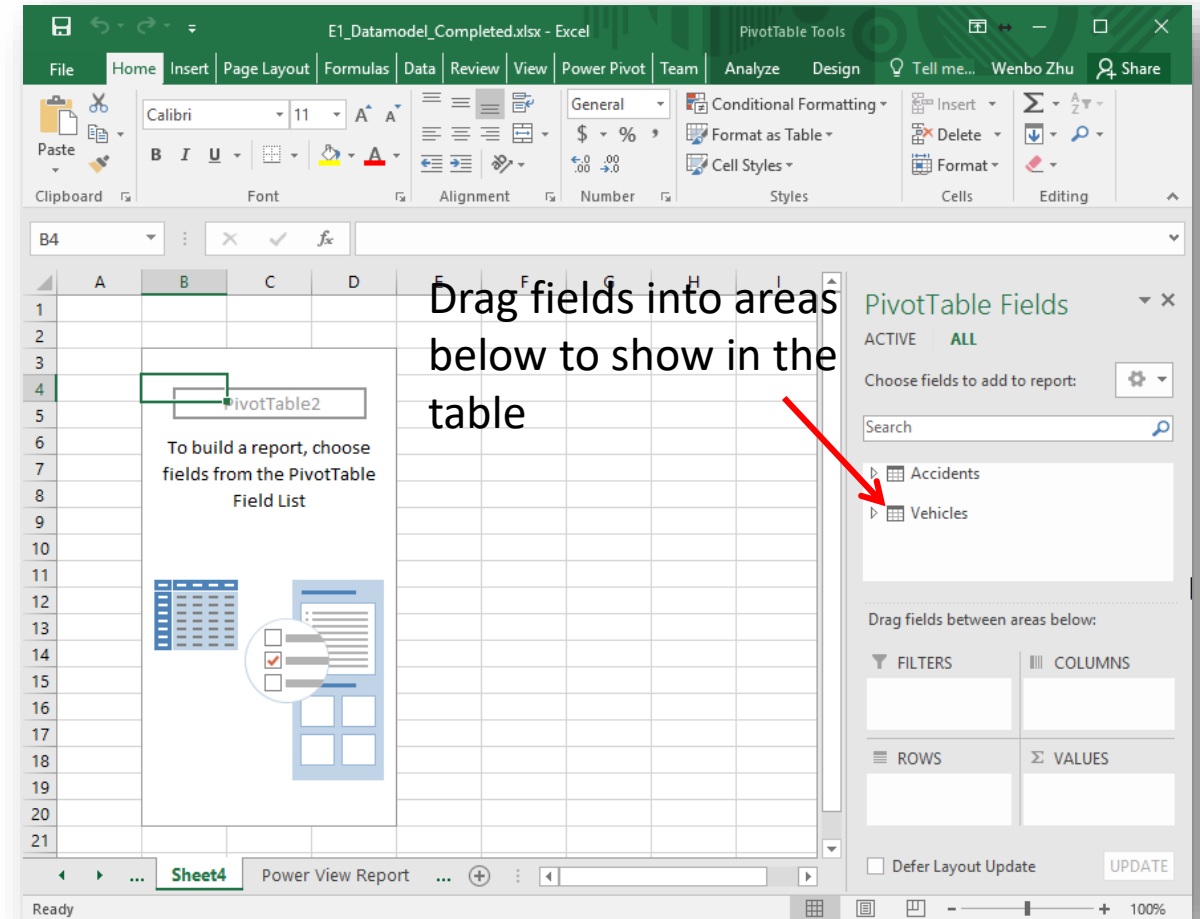
Step 5: Pivot Tables & Charts

- In Power Pivot, you are able to handle very large datasets (e.g., millions of rows) and analyze the data with low memory and high speed.

- Here we will look at two data analysis functions enabled in Power Pivot.
  - Pivot Table
  - Pivot Chart

- Although we are not using a very large dataset, the procedures are similar.

- In the Power Pivot window, click **PivotTable** under **Home** tab to create a PivotTable in Excel.

- Create the PivotTable in a new worksheet.
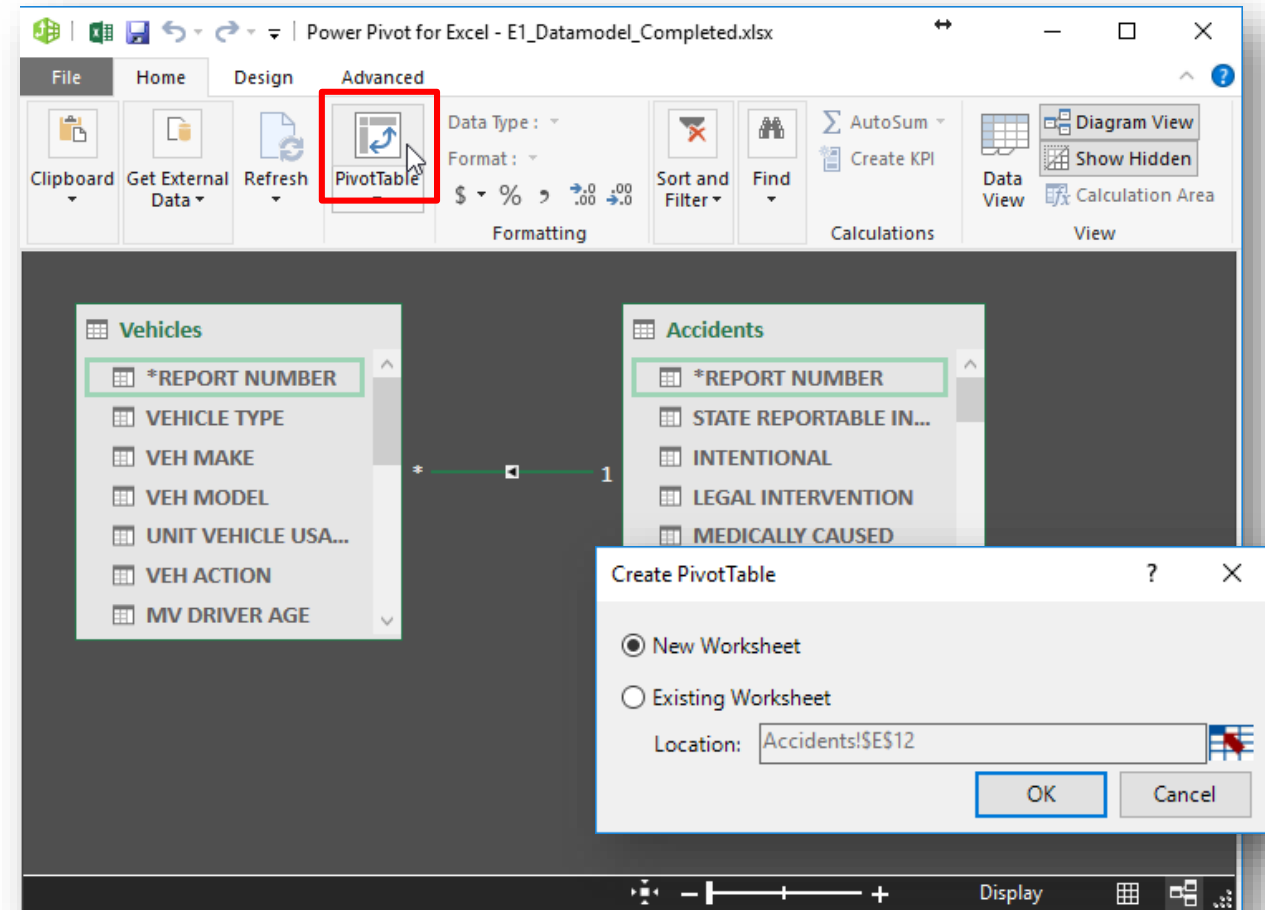
Step 5: Pivot Table & Charts

○ You will see a blank Pivot Table in the middle of a new Excel worksheet.

○ Click anywhere inside the table will open up the PivotTable Fields panel on the right side.

○ You can expand the tables and drag fields of interests into the four areas below.

○ Pivot Tables are a good way to summarize the datasets by using the four areas in the fields panel.

○ VALUES area controls the values to summarize (e.g., average, count, etc.).

○ The other three areas can break down your data into different levels

Drag fields into areas below to show in the table

Step 5: Pivot Table & Charts

◦ Here we look at a simple example:

◦ I want to count the number of accidents associated with a given combination of first vehicle involved driver gender and age, as well as roadway type.

◦ Thus, I need to:

◦ Use the vehicle role as a filter, to make sure I only look at the first vehicle.

◦ Use the Age and Driver gender as rows.

◦ Use the roadway type as columns (columns are rows are exchangeable).

◦ Use the COUNT aggregation function, and so the value that I am counting must be unique for any accident. I will use the state patrol accident report number.

Step 5: Pivot Table & Charts

◦ Drag "VehicleRole" from Vehicles table into FILTERS, "ROADWAY CHARACTERISTICS" from Accidents table into COLUMNS, "MV DRIVER AGE GROUP" and "MV DRIVER GENDER" from Vehicles table into ROWS, and "*REPORT NUMBER" from Vehicles table to VALUES as shown below.

◦ Make sure that you are using "*REPORT NUMBER" from Vehicles table rather than Accidents table (**why?**).
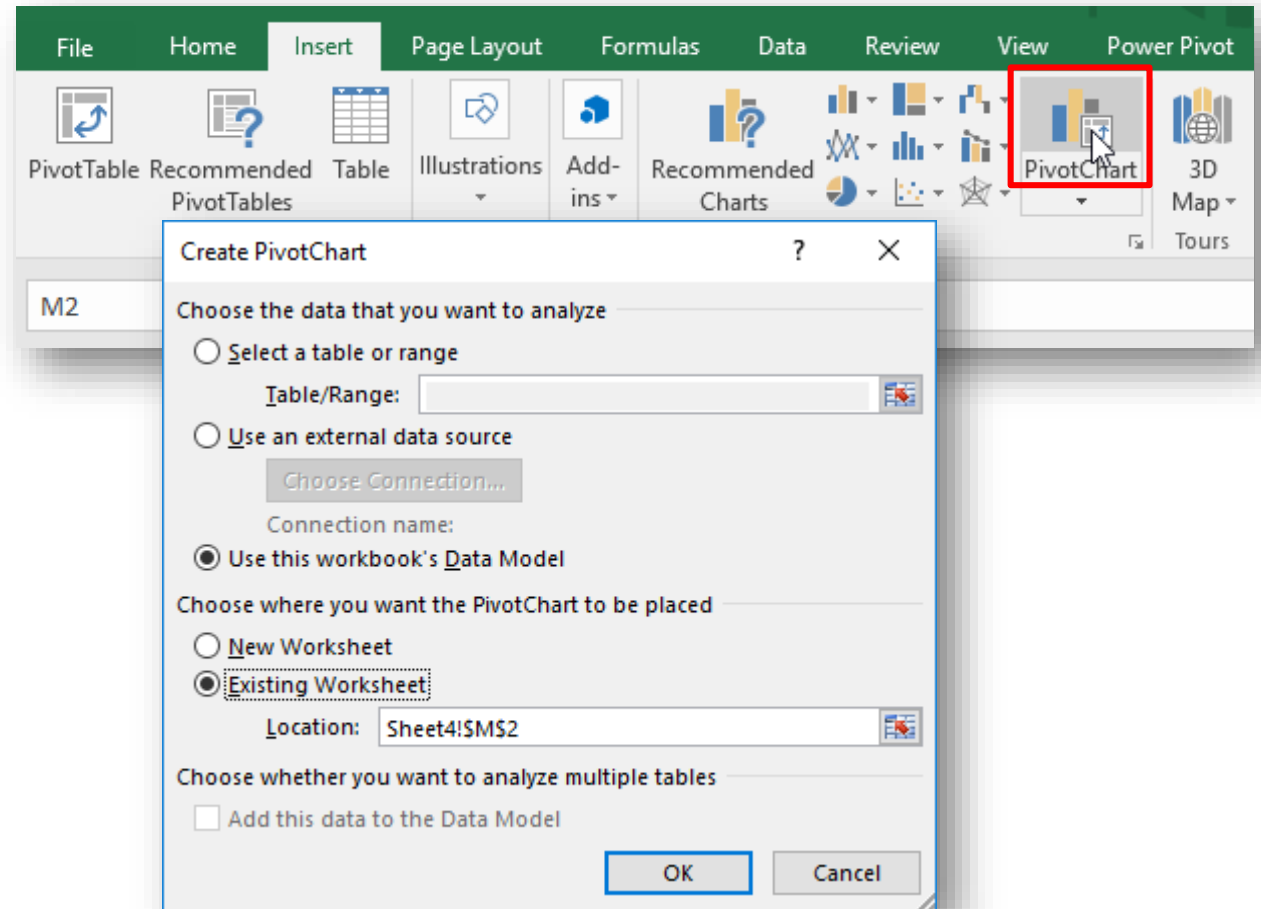
Step 5: Pivot Table & Charts

◦ You will get an accident count summary table as shown on the right.

◦ I changed "VehicleRole" filter to only show counts associated with primary vehicles.

◦ You can specify the selections of columns and rows in the dropdown lists to narrow down your results.

| VehicleRole	1							
Count of *REPORT NUMBER	**Column Labels**							
Row Labels	**Curve & Grade**	**Curve & Level**	**Curve in Sag**	**Straight & Grade**	**Straight & Level**	**Straight at Hillcr**	**Straight in Sag**	**Unknown**	**Grand Total**
⊟ < 20	6	4		9	74			1	94
Female	2	1		1	35			1	40
Male	4	3		8	38				53
unknown					1				1
⊟ 20-29	17	13	2	45	274	1	1	3	356
Female	9	5		16	97	1	1	1	130
Male	8	7	1	28	176			2	222
unknown		1	1	1	1				4
⊟ 30-39	17	12		25	154			3	211
Female	8	4		9	54			1	76
Male	9	8		16	98			2	133
unknown					2				2
⊟ 40-49	9	7	1	38	115		1	2	173
Female	3	1		13	38			1	56
Male	6	6	1	25	75		1	1	115
unknown					2				2
⊟ 50-59	8	3		15	91	2	2	3	124
Female	2	1		4	31	1	2	1	42
Male	6	2		11	59	1		2	81
unknown					1				1
⊟ 60-69	6	5		12	47		1		71
Female	4	2		5	18				29
Male	2	3		6	29		1		41
unknown					1				1
⊟ 70-79	1			3	18		1		23
Female				2	7				9
Male	1			1	11		1		14
⊟ 80-89				2	4				6

Step 5: Pivot Table & Charts

◦ You can also create Pivot Charts to summarize your data graphically.

◦ Click anywhere outside the Pivot Table. Under Insert, click Pivot Chart.

◦ In the prompted window, keep all options as default to create a PivotChart on the current worksheet.

Step 5: Pivot Table & Charts

◦ The Pivot Chart has a similar fields panel.

◦ Drag fields to the areas below to make a chart corresponding to the Pivot table as shown.

◦ The Pivot chart may have too many columns. Try narrow down your selections (e.g., only show one road type / gender / age group).
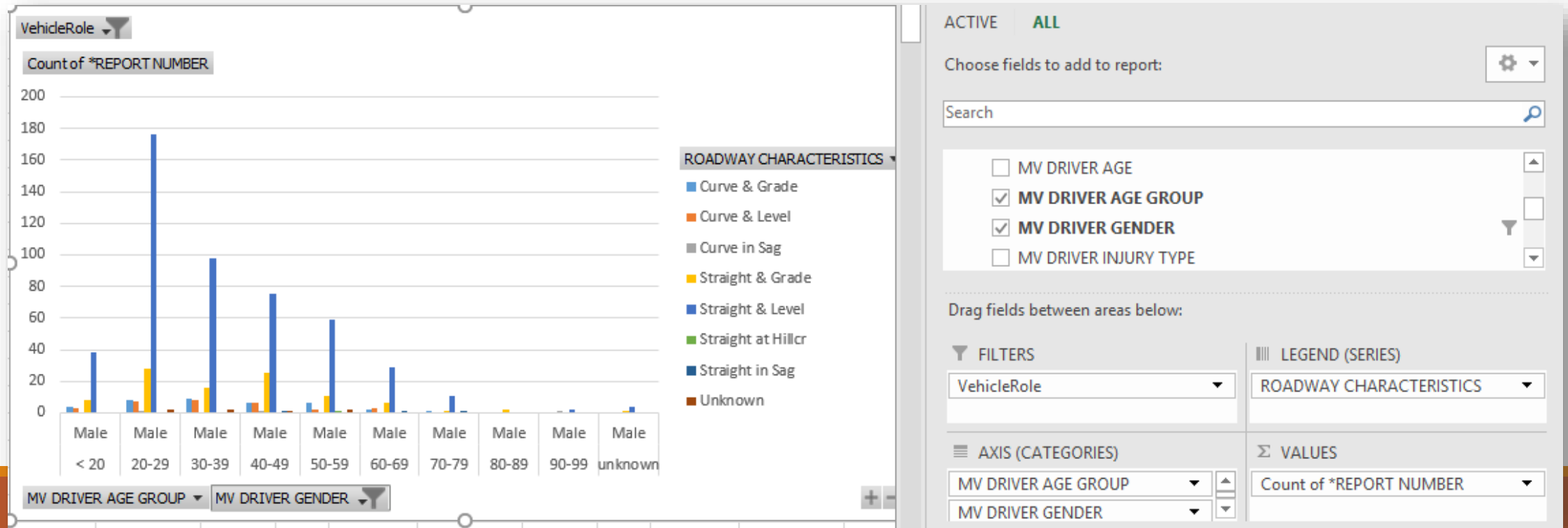
Step 6: Generate Power View Chart

◦ We just did two quick examples to summarize and visualize data in Power Pivot. Note that the key features of Power Pivot are big data support and data management functions. A typical data modeling procedure in Excel is to build and manage your data model in Power Pivot, and then visualize the data in Power View.

◦ In the rest of the exercise, we will generate more summary plots to describe the data using Power View.

◦ In the Excel main window, navigate to the Power View report you created.

◦ You can close your Power Pivot window now as the relationship has already been created.

Step 6: Generate Power View Chart

- ◦ Your Power View Report may have included a default table. Delete it and start from a blank report.

- ◦ In the Power View Fields panel on the right, select the "*REPORT NUMBER", "MV DRIVER INJURY TYPE", and "MV DRIVER RESTRAINT" fields from the Vehicles table as shown.

- ◦ With all fields added, click on the drop down arrow next to "*REPORT NUMBER" and set the aggregation to **Count (Not Blank)** as shown. This way, we can count the number of drivers associated with each distinct injury type.

Step 6: Generate Power View Chart

◦ Click inside the table that has been created, and click **Design** → **Other Chart** → **Pie** as shown on the right.

◦ Drag the "MV DRIVER RESTRAINT" field to the **COLOR** area below "MV DRIVER INJURY TYPE"

◦ Resize the resulting pie chart to make sure that all of the fields and text are visible. You should have something like what is shown on the right.
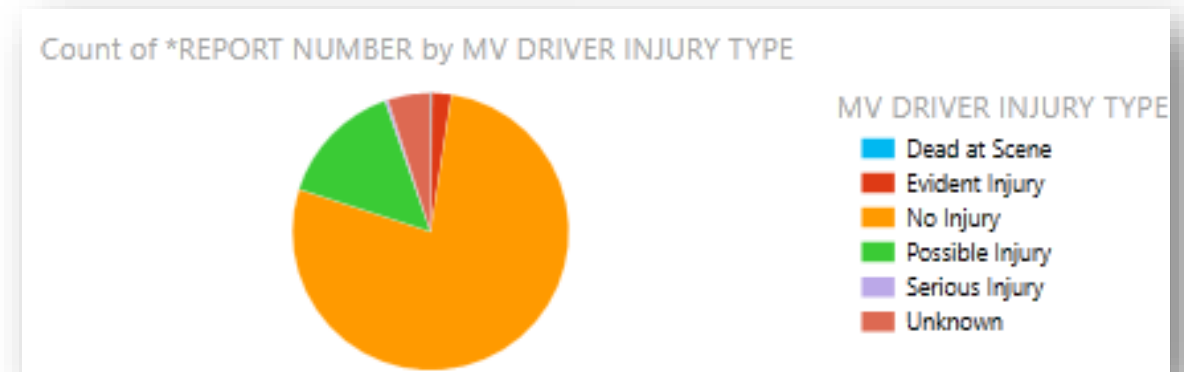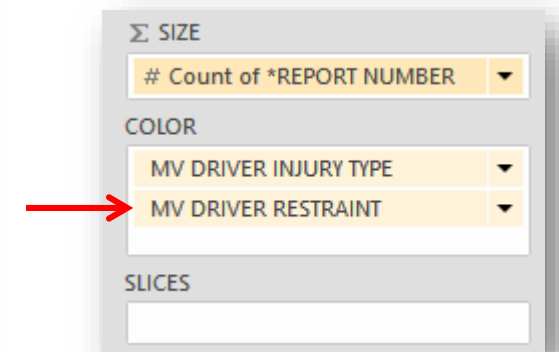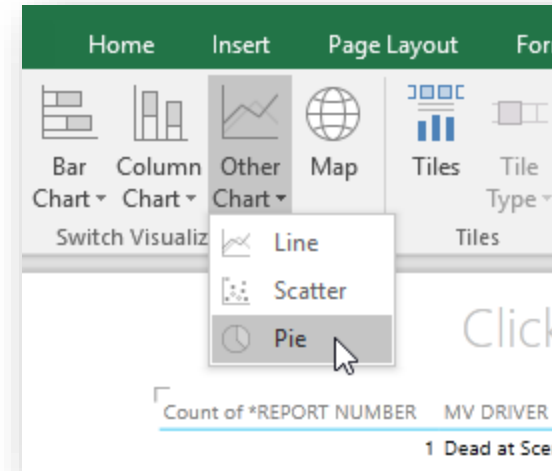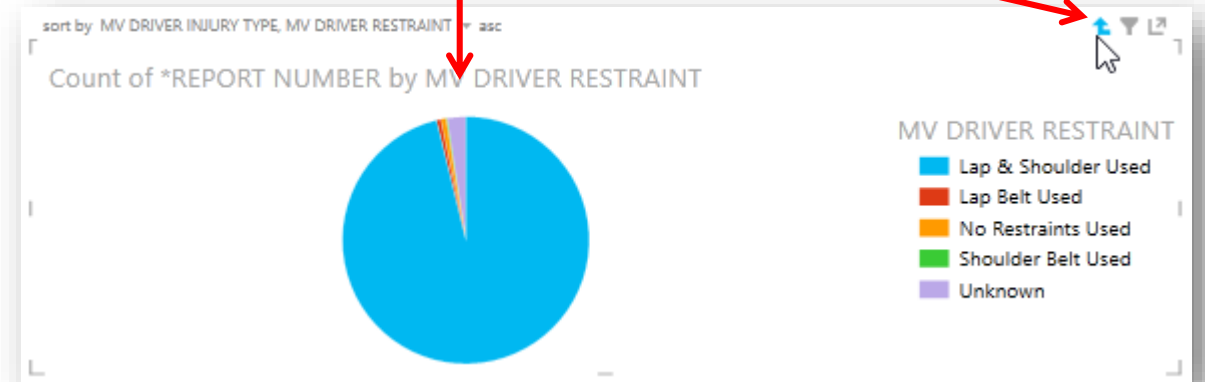
Step 6: Generate Power View Chart

◦ Note that the pie chart only shows the breakdown of accidents by injury type. This is because the "MV DRIVER RESTRAINT" field is on the second level.

◦ Double click any slice of the pie chart to view the accident breakdown by driver restraint conditions for the corresponding injury type.
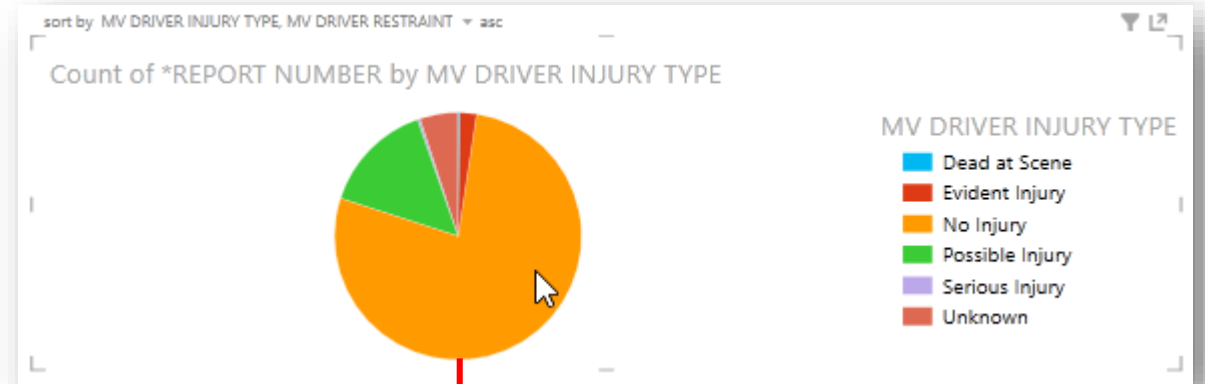

Double click any slice to view second-level breakdown

Click here to move up

Step 6: Generate Power View Chart

◦ Note that the pie chart has not taken advantage of the relationship between the two tables. We will now use the relationship to count the number of drivers in each age group, separated by accident time of day.

◦ Click anywhere blank in the Power View report. Start by adding "AMPM" from the Accident table, and then add "*REPORT NUMBER" and "MV DRIVER AGE GROUP" from the Vehicles table as shown.

◦ You will need to change the ordering of the fields, you can do this by dragging a field from its position in the FIELDS list to a new position. Match the ordering shown on the right.
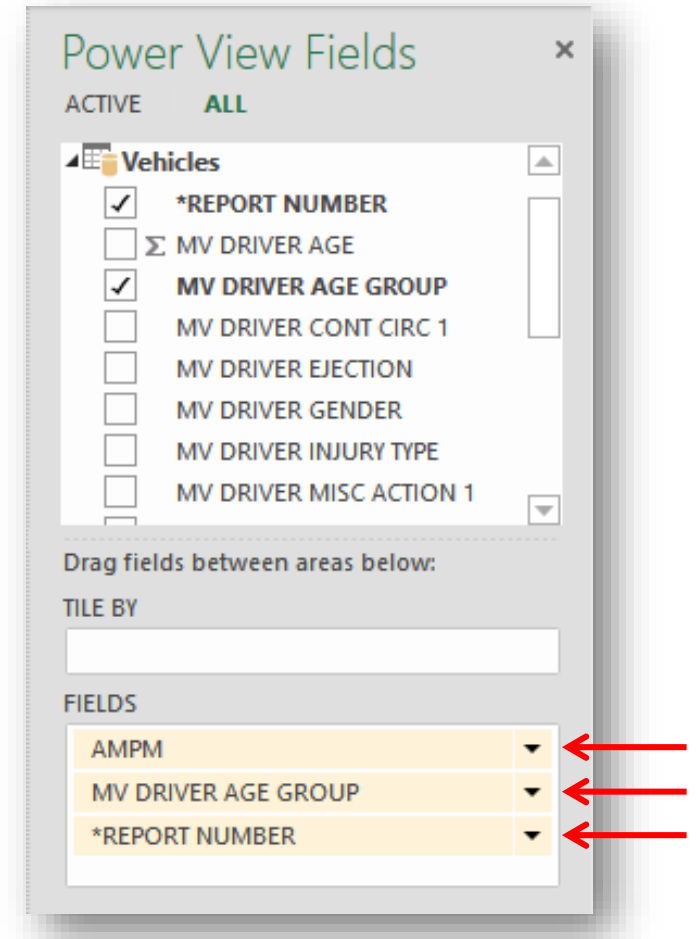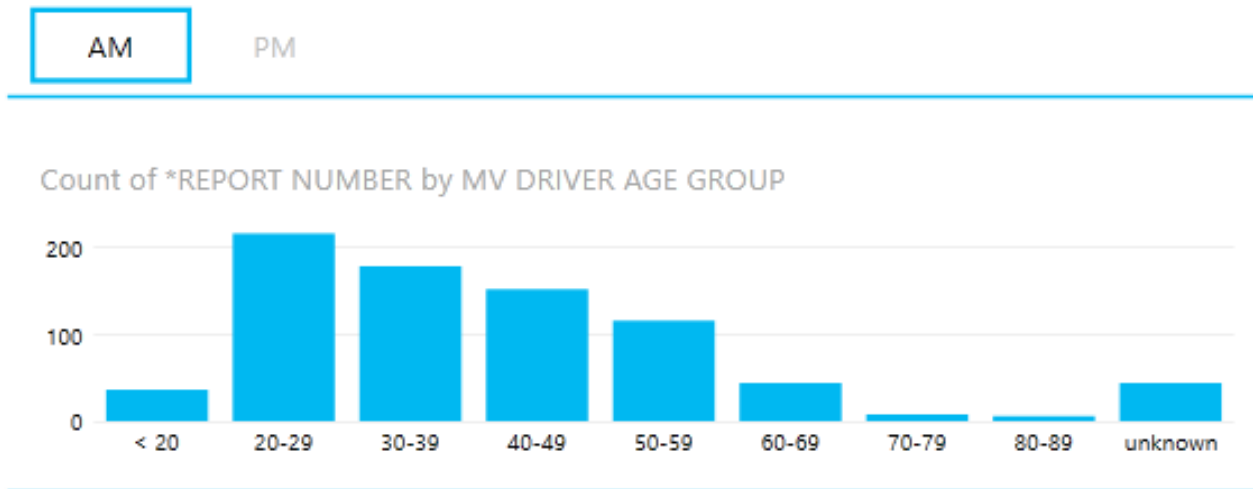
Step 6: Generate Power View Chart

◦ Once again, you need to change the aggregation on the "*REPORT NUMBER" field to **Count (Not Blank)** as shown.

◦ Now you have a table describing the driver count for each distinct combination of AM, PM, and driver age group.

◦ Again after clicking inside the table, click **Design → Column Chart → Stacked Column** as shown on the lower right.
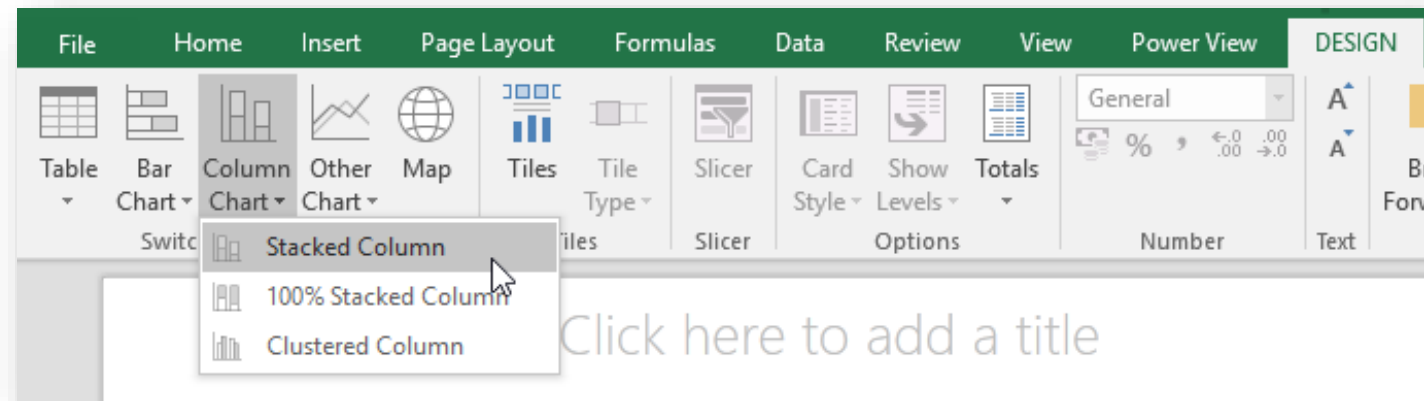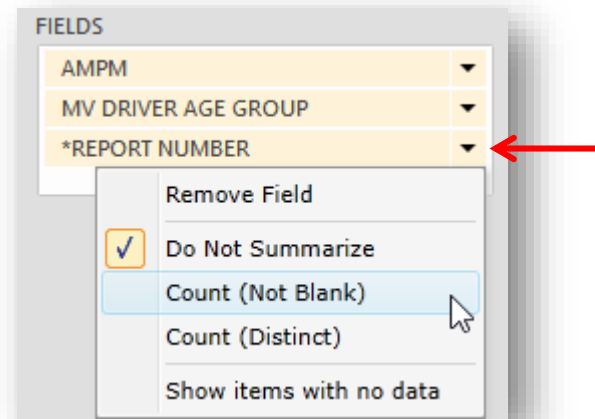
Step 6: Generate Power View Chart

○ You plot will probably look wrong, change the structure by moving the AMPM field to TILE BY, the Count of REPORT NUMBER to Values, and MV DRIVER AGE GROUP to AXIS as shown on the right.

○ The result should look similar to what is shown below, after some resizing. Note you can click on AM or PM to view a summary of each time period. Try clicking on one or more of the vertical bars.
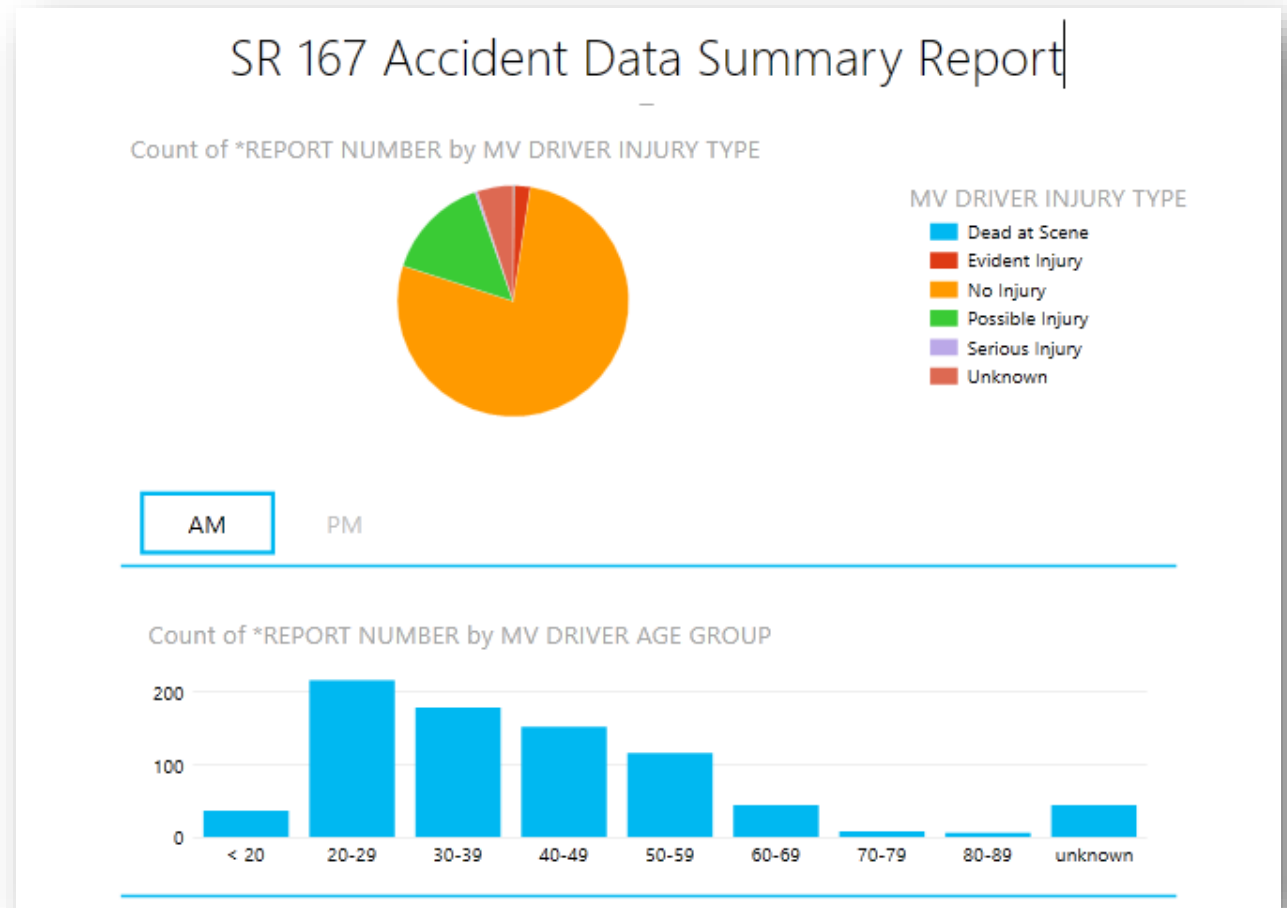
Step 6: Generate Power View Chart

◦ Your report should look something like what is shown on the right.

◦ You can click on any of the fields to change the view, and your selection will be reflected in the rest of the report.

◦ For example, try clicking in the "No Injury" section of the pie chart, or on one of the age group bars in the bar chart.

SR 167 Accident Data Summary Report

Count of *REPORT NUMBER by MV DRIVER INJURY TYPE

MV DRIVER INJURY TYPE
- Dead at Scene
- Evident Injury
- No Injury
- Possible Injury
- Serious Injury
- Unknown

AM   PM

Count of *REPORT NUMBER by MV DRIVER AGE GROUP

< 20   20-29   30-39   40-49   50-59   60-69   70-79   80-89   unknown

Step 6: Generate Power View Chart

◦ We need to add another Power View Report sheet before proceeding. Do this as we did the first time, by clicking **Insert → Power View**.

◦ Let us look at the accident distribution along the route. Add the "MILE POST WHOLE" and "QUARTER" fields from the Accidents table, and the "*REPORT NUMBER" and "VEH COMP DIR TO" fields from the Vehicles table. The "VEH COMP DIR TO" field will serve as an indicator of what direction of travel the accident took place on.

◦ The aggregation of "MILEPOST WHOLE", "QUARTER", and "VEH COMP DIR TO" should be set to **Do Not Summarize**. The aggregation on "*REPORT NUMBER" should be set to **Count (Not Blank)**. Order them as shown on the right.
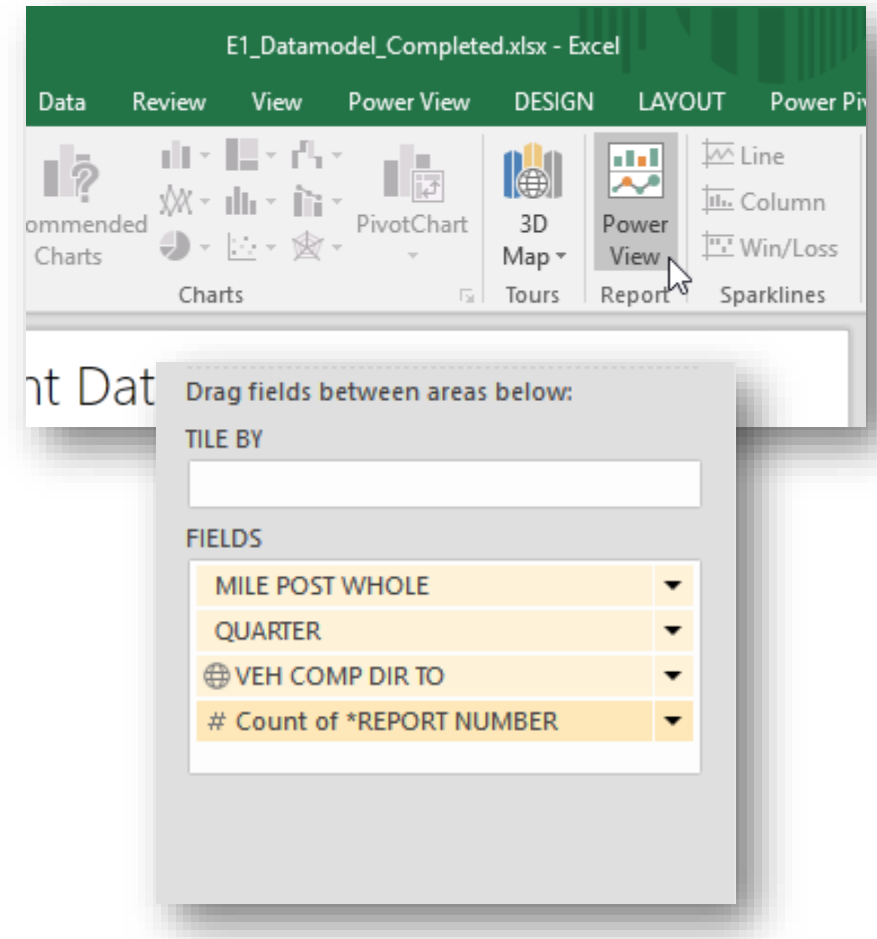
Step 7: Additional Plots and Summaries

◦ Now insert a line plot by first clicking inside the newly created table, and then clicking **Design → Other Chart → Line** as shown.

◦ Again, the result will probably look weird, and will require some changes before we get what we want. The next slide will illustrate this.
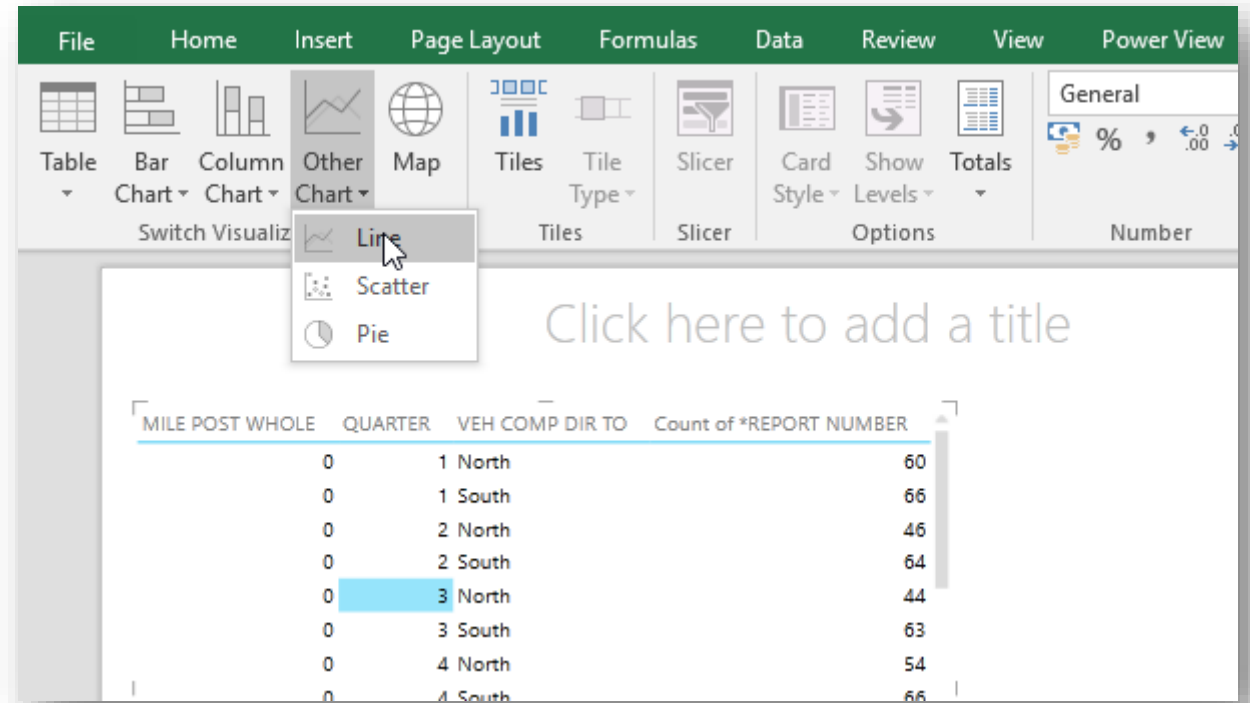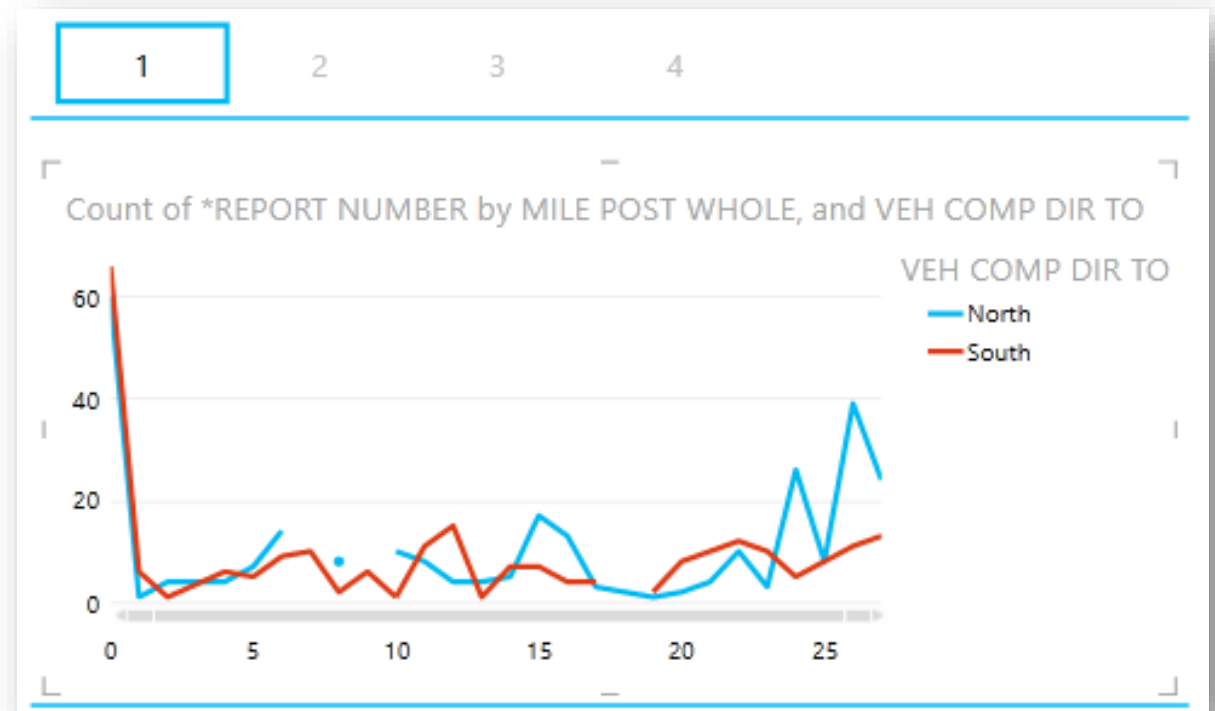
Step 7: Additional Plots and Summaries

◦ You should move the fields around such that "QUARTER" is the **Tile By** field, "Count of *REPORT NUMBER" is the **Values** field, "VEH COMP DIR TO" is the **Legend** field, and "MILE POST WHOLE" is the **Axis** field as shown on the right.

◦ Depending on how these fields are assigned by default, Power View may not allow to assign "QUARTER" as the **Tile by** field initially. You need to first make sure that "MILE POST WHOLE" is the **Axis** field, and that there are no fields assigned to vertical or horizontal multiples.
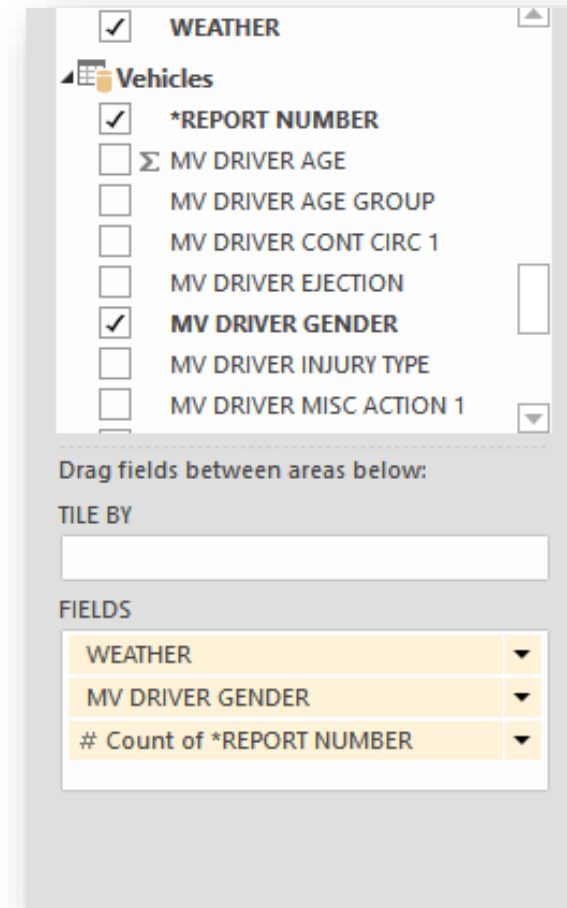
Step 7: Additional Plots and Summaries

◦ You should get something similar to what is shown on the right. Note that you can click the numbers at the top of the plot to change the quarter, and view the accident counts for each milepost and quarter (that is, quarter of the year).

Step 7: Additional Plots and Summaries

◦ Finally, let us look at accident count by gender and weather conditions. Add a new table to your Power View Report by clicking in an empty space in the report and checking the boxes next to "MV DRIVER GENDER" and "*REPORT NUMBER" in the Vehicles table, and "WEATHER" in the Accidents table.

◦ Set the aggregation type for "*REPORT NUMBER" as **Count (Not Blank)**.

◦ You should get something similar to what is shown on the right.

◦ Now, after clicking inside your newly created table, click **Design → Column Chart → Stacked Column**.

Step 7: Additional Plots and Summaries

◦ If you then arrange your fields as shown on the right, you should get a stacked column chart similar to what is shown below.
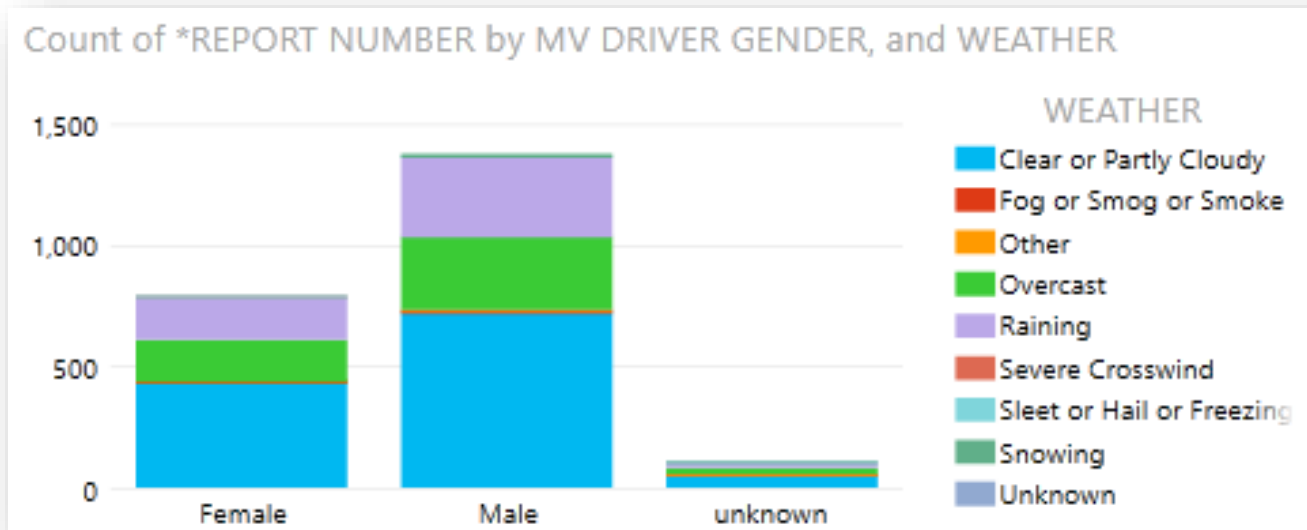
Step 7: Additional Plots and Summaries

○ A Filters menu should be available on the right of the Power View Report as shown on the right. If not, you can make it visible by clicking on the filter icon on the top right corner of your plot.
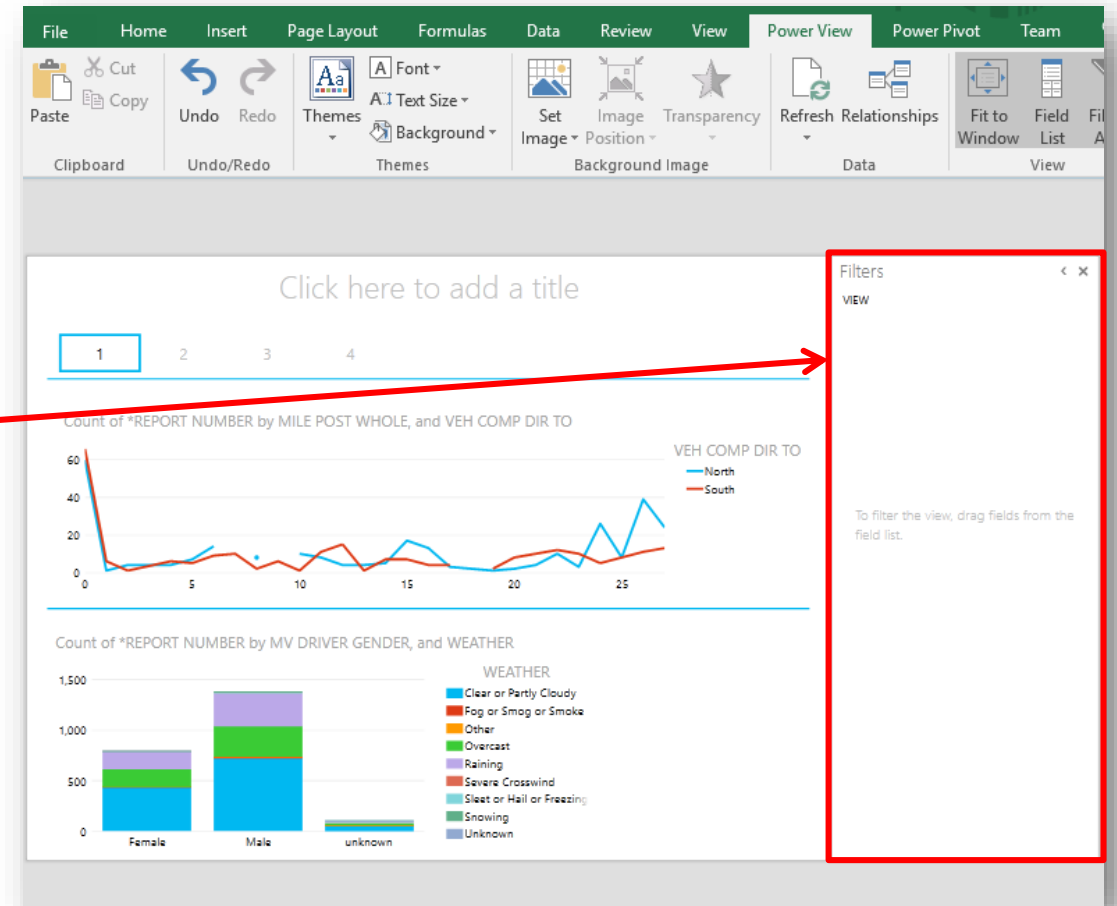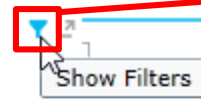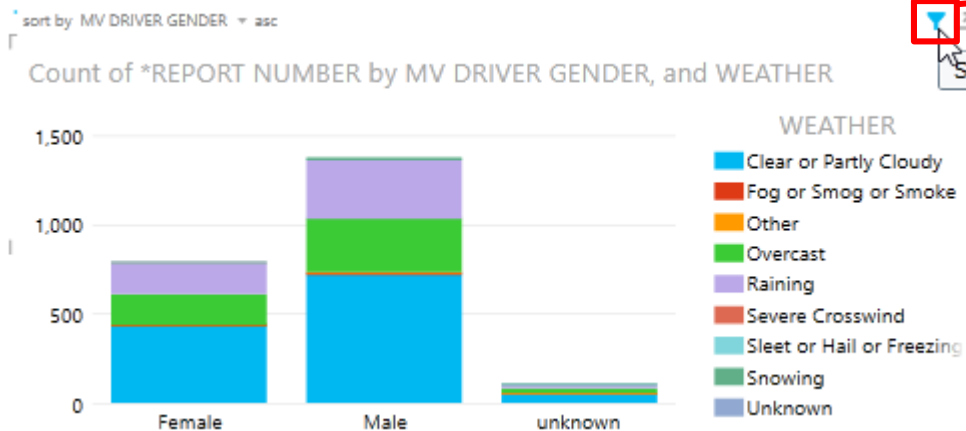
Step 7: Additional Plots and Summaries

◦ Most of these weather types only appear a couple times in the dataset, so we will filter out the minor ones. Click inside the chart, and then click CHART under the filter menu as shown below. You can expand the options under WEATHER, and select only the weather conditions corresponding to at least 10 accidents as shown.
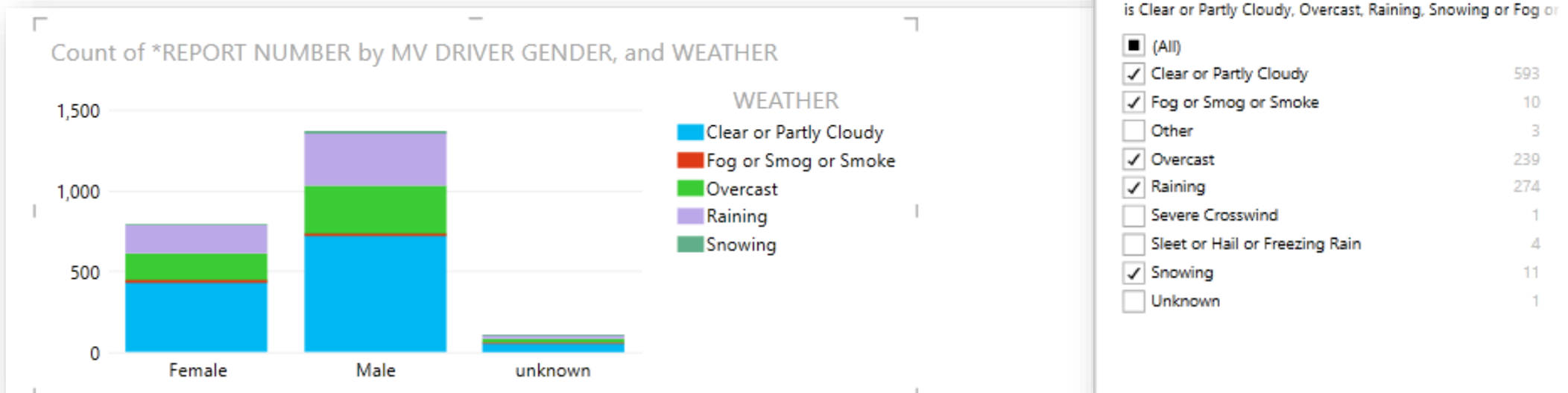
Step 8: What Else?

◦ Do some exploring and make something interesting.

◦ Note: you can utilize a number of data sources in Power View, and you can get even more capability by using Power Pivot to define more complex data models. These tools are more powerful when used together.