



# SAS University Edition Challenge

Day 5 of 5 – Friday, 20th July 2018

## Challenge Overview

You and your friend (let's call him) Kenny, have decided to start an online website that aggregates movie and TV show data streamed by Netflix.

Over the last few days, you have conducted preliminary analysis on the individual data sets, appended them into one large dataset, and finally, impressed Kenny with some summary statistics. Kenny has advised you that the database is just about ready, but he'd like to see some graphical analysis and would like to know which Netflix titles have the highest rating.

You agree and get started!

## Summary of Skills Demonstrated

- Use box plots to visualize summary statistics for groups.
- Be able to Query the data.

## Submission Details

Submissions close **Monday, 23<sup>rd</sup> July 2018, 12:00pm.**

Submit your answers [here](#).

Be sure to use your SAS account when you submit your solution. If you don't use your SAS account, we can't put your entry in the draw!

If you haven't registered yet, it's not too late. Click [here](#) to register now.

Also, make sure to save your answers and tasks/code somewhere safe as a backup.

## SAS University Edition

You'll need to download and install SAS University Edition to complete this challenge. Use the links below to install and set up SAS University Edition.

SAS University Edition download link:

[https://www.sas.com/en\\_au/software/university-edition/download-software.html](https://www.sas.com/en_au/software/university-edition/download-software.html)

How do I create a shared folder in VirtualBox?

[https://support.sas.com/software/products/university-edition/faq/shared\\_folder\\_virtualbox.htm](https://support.sas.com/software/products/university-edition/faq/shared_folder_virtualbox.htm)



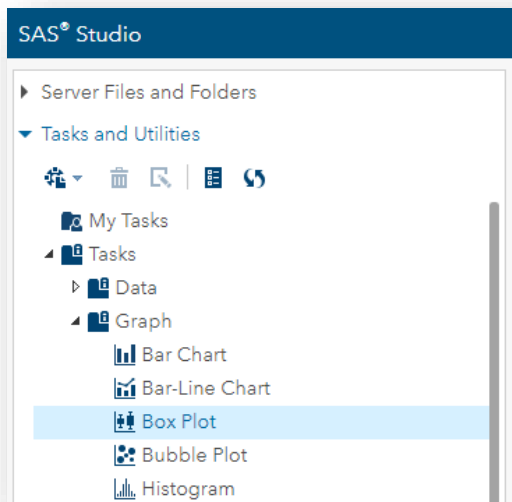
## Guided Exercises

### Before we start

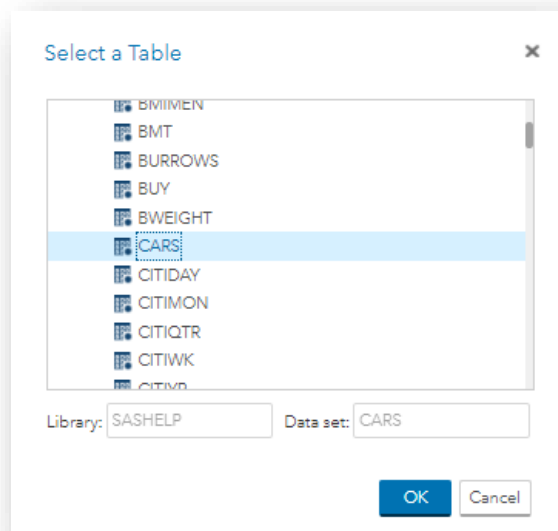
Be sure to have completed the challenges from previous days.

### Generate Box Plots

1. In the left pane, expand **Tasks and Utilities**, then expand **Graph**.
2. Double-click the **Box Plot** task to open it.

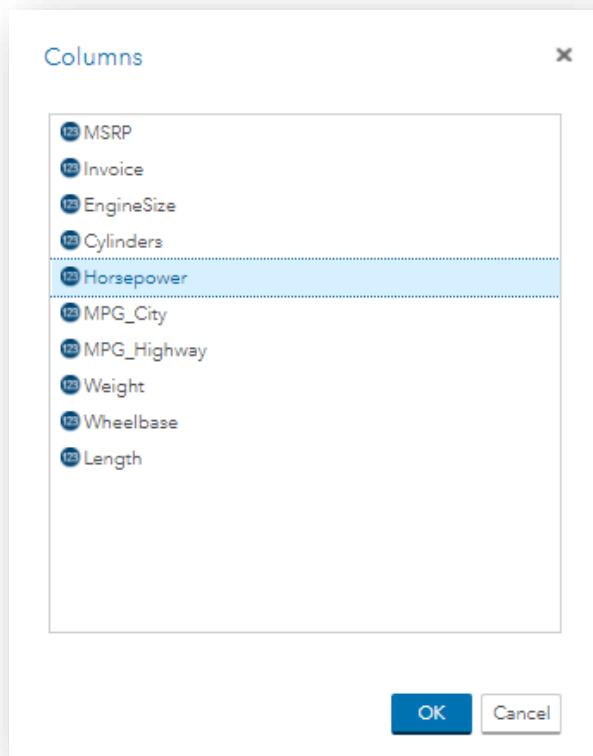


3. Under the **Data** heading, select the **CARS** data set from the **SASHELP** library.

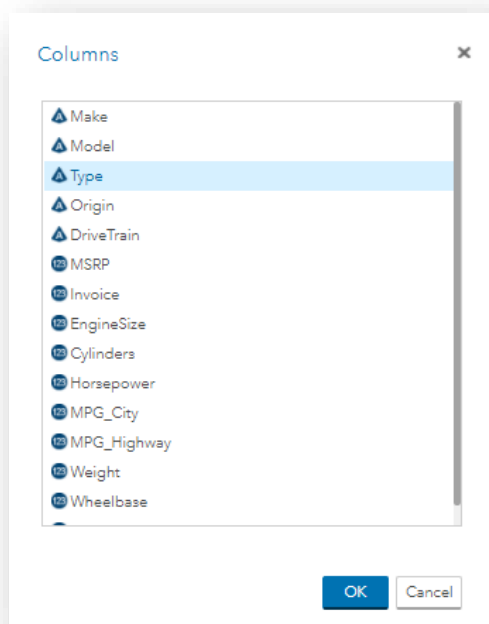





4. Under **Roles**, click the '+' icon next to **Analysis variable**. Select **Horsepower** and click **OK**.

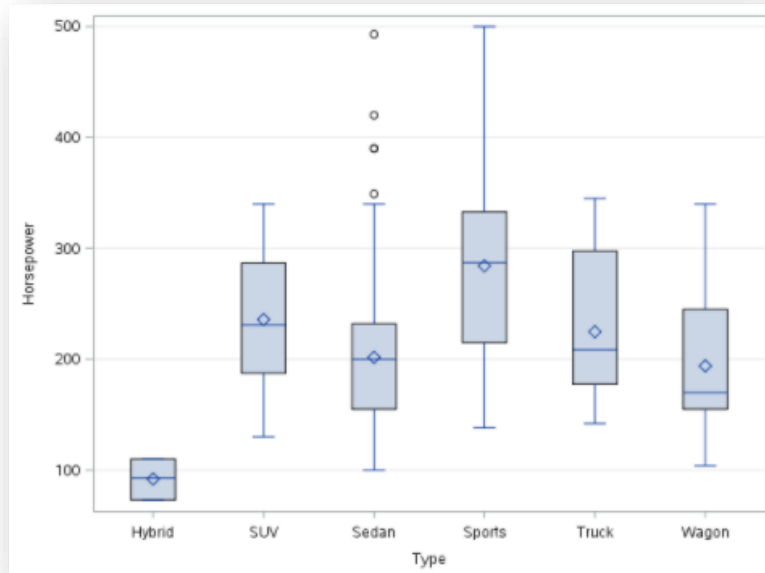


5. Under **Roles**, click the '+' icon next to **Category**. Select **Type** and click **OK**.





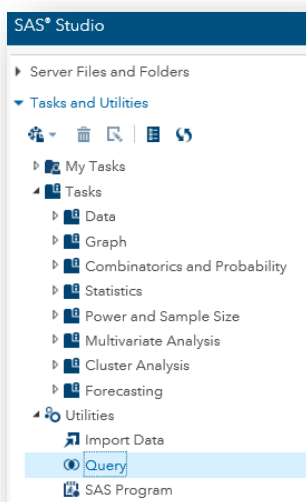
6. Click **Run** 
7. Under **Results** tab you will see the following box plots.



8. Hover your cursor over the **SUV** box plot to see the **Summary Statistics**.

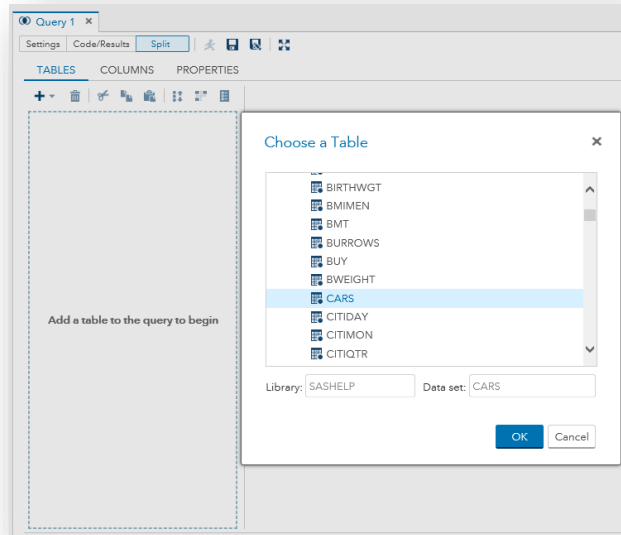
## Query the data

1. Navigate to **Tasks and Utilities** in the left pane. Expand the **Utilities** menu.
2. Double-Click the **Query** task to open it.

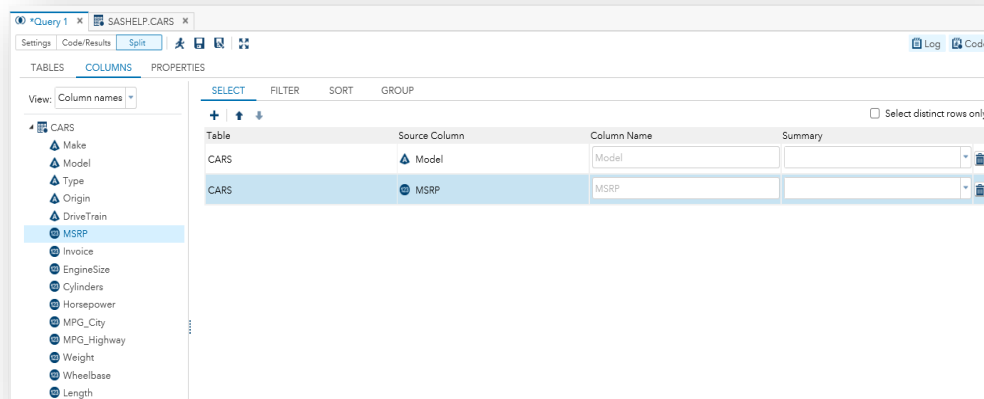




3. Under the **Tables** heading, click the '+' and select **Table**. Then select the **CARS** data set from the **SASHELP** library. Click **OK**.

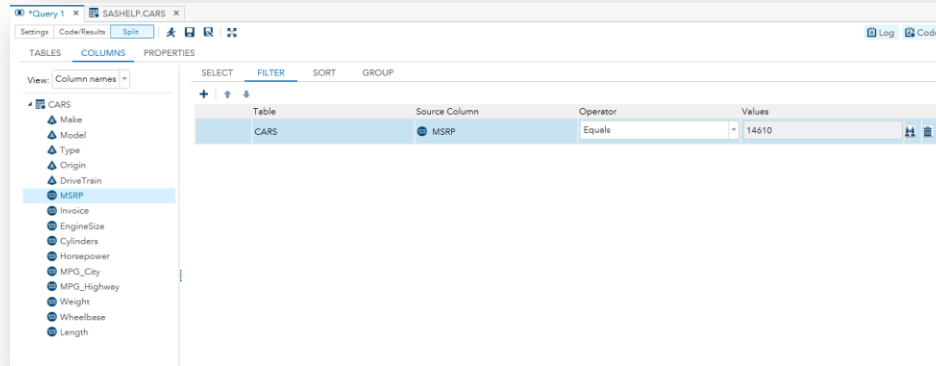


4. Under the **Columns** heading, select the **Select** Tab. Expand the column names of the **CARS** dataset on the left-hand side. Then drag and drop **Model** and **MSRP** on to the Query pane.





- Under the **Columns** heading, select the **Filter** Tab. Drag and drop **MSRP** on to the Query pane. For the **Operator** field select **Equals**. Then Click the *binoculars* icon for the **Values** field. Type **14610** under **Select Value** for 'Value:'.



- Click **Run**

On the bottom pane under the **OUTPUT DATA** Tab you will be able to see that the model of a car with a MSRP price of \$14,610 is a Cavalier 2dr.

To learn more about creating graphs, check out the [SAS video portal](#).

End of Guided Exercises



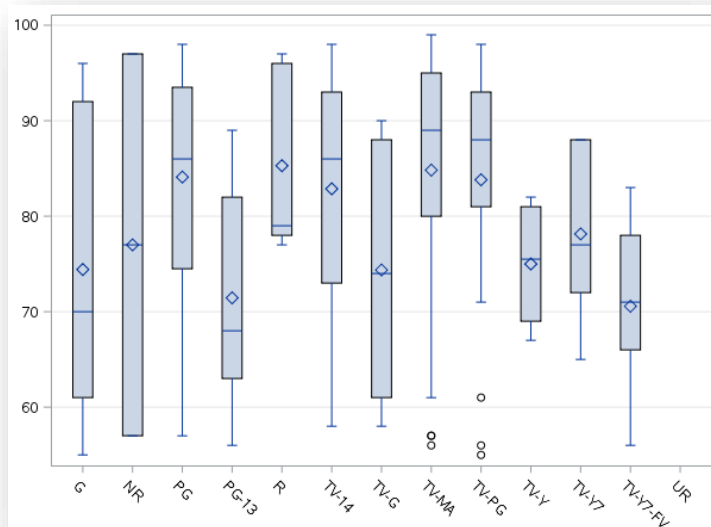
## Challenge Exercises

Submit your answers [here](#). Answer all 4 questions correctly to gain an entry into the prize draw.

This challenge requires information from the previous challenges. Be sure to complete the other challenges before completing this one.

It's a good idea to save your tasks in an accessible location once complete. If anything happens to your data, you can always rerun the tasks.

Re-create the following graph using the **NETFLIX\_FINAL** data set from the **NETFLIX** library:



Question 1: What **Category variable** and **Analysis variable** were used to create this graph?

Question 2: How many outliers are there in these groups?

Question 3: Using the **Tooltip**, what is the mean for 'PG' titles?

Using the Query task with the **NETFLIX\_FINAL** data set answer the following question:

Question 4: Which Netflix **TITLE** has a **user\_rating\_score** of 99?

Don't forget to visit our Facebook page for the previous challenges!



<https://www.facebook.com/SASAustNZ>