

Course Title

## Big Data Emerging Technologies

### ❖ Modules

1. Big Data Rankings & Products
2. Big Data & Hadoop
3. Spark
4. Spark ML & Streaming
5. Storm
6. IBM SPSS Statistics Project

## Big Data IBM SPSS Statistics Project

## IBM SPSS Statistics Project

### ❖ Corporate Data Analysis Experience

- Many corporations provide a free trial period of 2 weeks ~1 month to experience their data analysis system
  - IBM SPSS Statistics, SAP S/4 HANA, Splunk Enterprise, Oracle DBMS, Microsoft Azure, etc.
- Now lets sign up to create an account and get some experience in using IBM's advanced corporate data analysis technology

## IBM SPSS Statistics Project

### ❖ IBM SPSS Statistics



- IBM's Rankings
  - 1<sup>st</sup> in Big Data Corporations
  - 4<sup>th</sup> in Big Data Software
  - 2<sup>nd</sup> in Big Data Hardware
  - 1<sup>st</sup> in Big Data Professional Services
- IBM SPSS
  - Initially release in 1968 by SPSS Inc., which was acquired by IBM in 2009

## IBM SPSS Statistics Project

### ❖ IBM SPSS Statistics

- Provides statistical analysis of logical batched and non-batched data based on descriptive statistics, regression, advanced statistics, etc.
- Various solutions are provided for Spark, SQL, text analytics for product/service integration and Big Data analysis
- SPSS Syntax enhancements using R and Python



## IBM SPSS Statistics Project

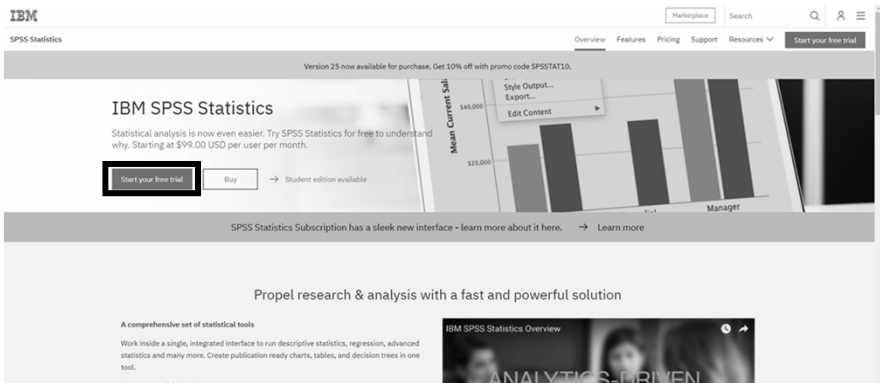
### ❖ Project 1: IBM SPSS Statistics Tutorial

- IBM SPSS Project Objective
  - Gain insight of advanced data analysis using IBM SPSS
- IBM SPSS's functions
  - Provides comprehensive set of statistical tools
  - Integration with Open Source software
  - Easy to use statistical analysis

# IBM SPSS Statistics Project

## ❖ Project 1: IBM SPSS Statistics Subscription

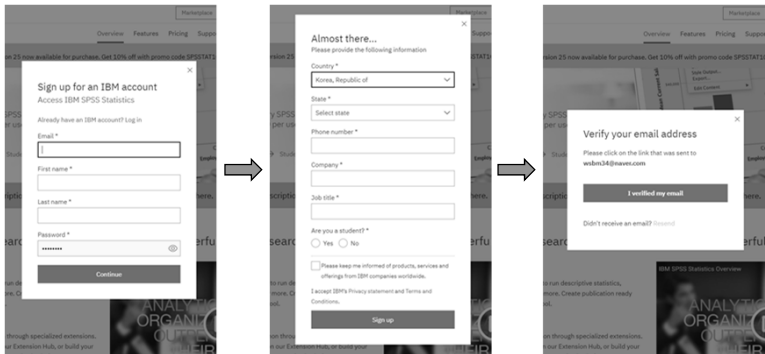
- Step 1: Make IBM account and start free trial
  - ➔ [https://www.ibm.com/products/spss-statistics?lnk=STW\\_US\\_MAST\\_L1\\_TL&lnk2=learn\\_SPSSstatSub&hpmht=a](https://www.ibm.com/products/spss-statistics?lnk=STW_US_MAST_L1_TL&lnk2=learn_SPSSstatSub&hpmht=a)



# IBM SPSS Statistics Project

## ❖ Project 1: IBM SPSS Statistics Subscription

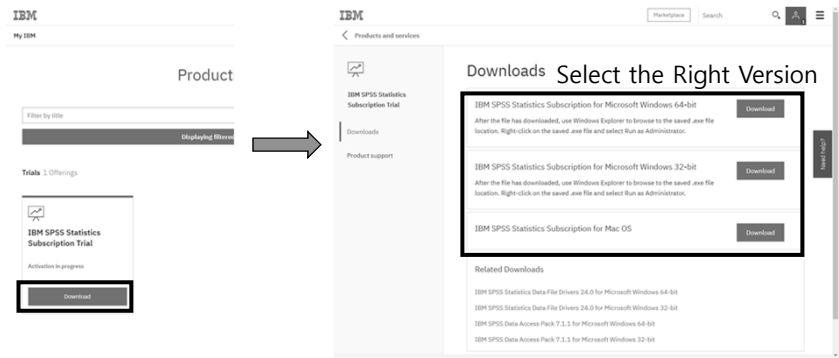
- Step 1-1: Write your information



# IBM SPSS Statistics Project

## ❖ Project 1: IBM SPSS Statistics Subscription

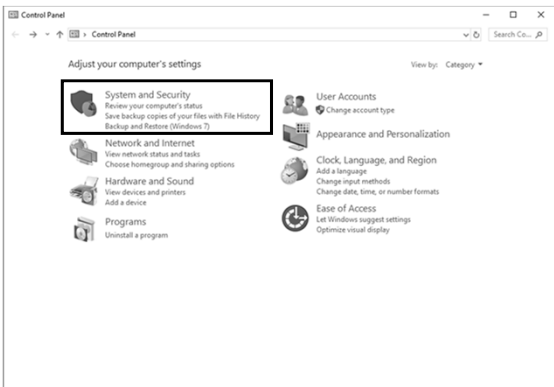
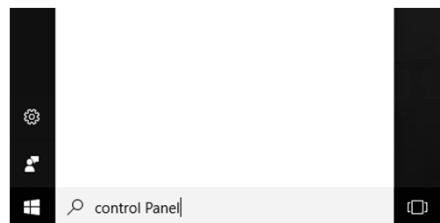
- Step 1-2: Download the Installation File



# IBM SPSS Statistics Project

## ❖ How to identify this on my computer

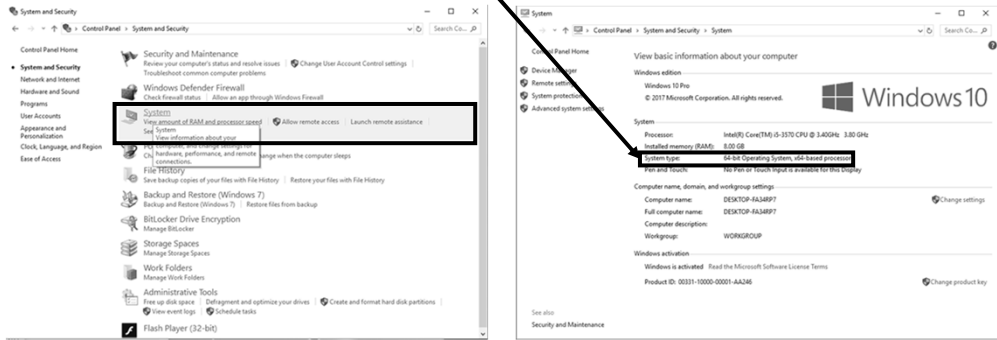
1. Open the “Control Panel”
2. Click “System and Security”



# IBM SPSS Statistics Project

## ❖ How to identify this on my computer

3. In "System and Security" click on "System"
4. Find "System type" ➔ **Windows 32 bit or 64 bit info is here!**



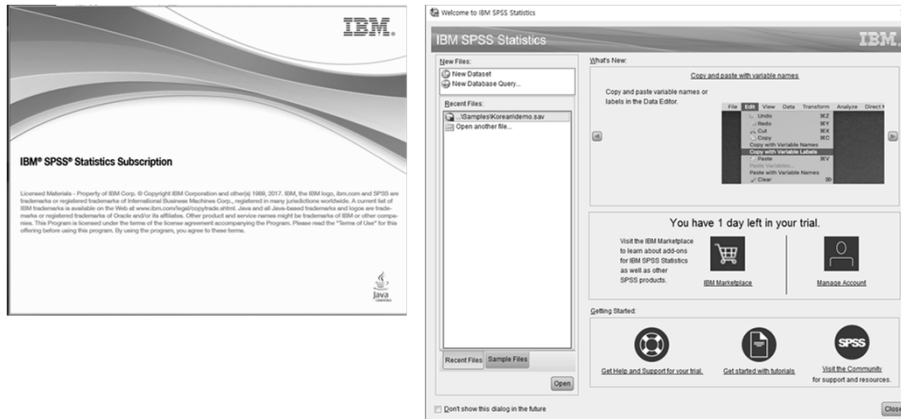
# IBM SPSS Statistics Project

## ❖ Project 1: IBM SPSS Statistics Subscription

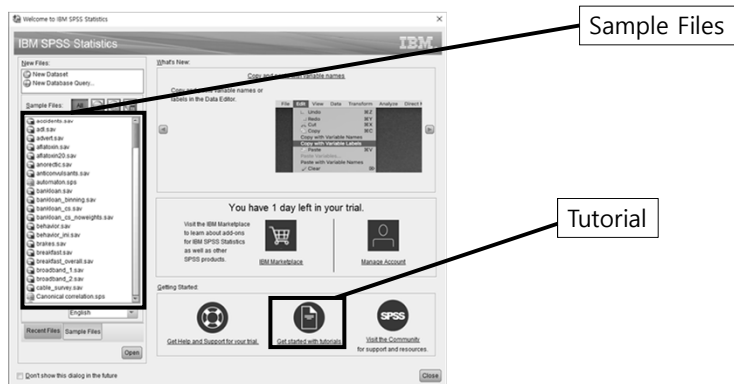
- Step 1-3: Download install file



- Step 2: Operating IBM SPSS Statistics



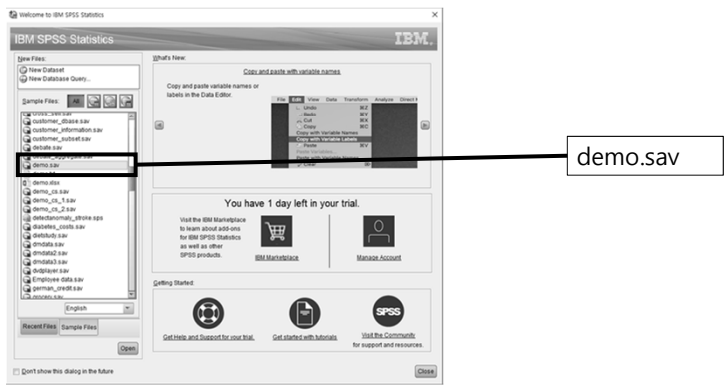
- Step 2: Operating IBM SPSS Statistics



# IBM SPSS Statistics Project

## ❖ Project 1: IBM SPSS Statistics Subscription

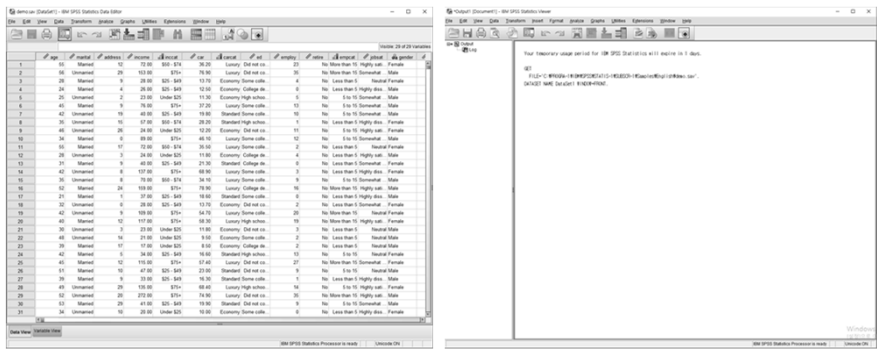
- Step 3: Open Data Sample Files “demo.sav”



# IBM SPSS Statistics Project

## ❖ Project 1: IBM SPSS Statistics Subscription

- Step 3: Windows “Data Editor” & “Output Viewer”

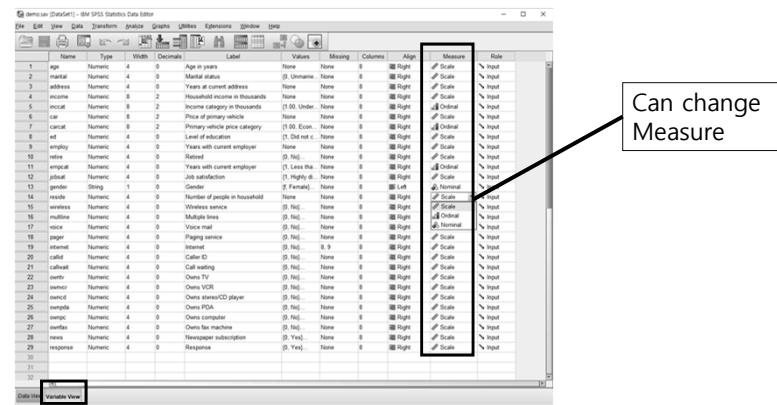




# IBM SPSS Statistics Project

## ❖ Project 1: IBM SPSS Statistics Subscription

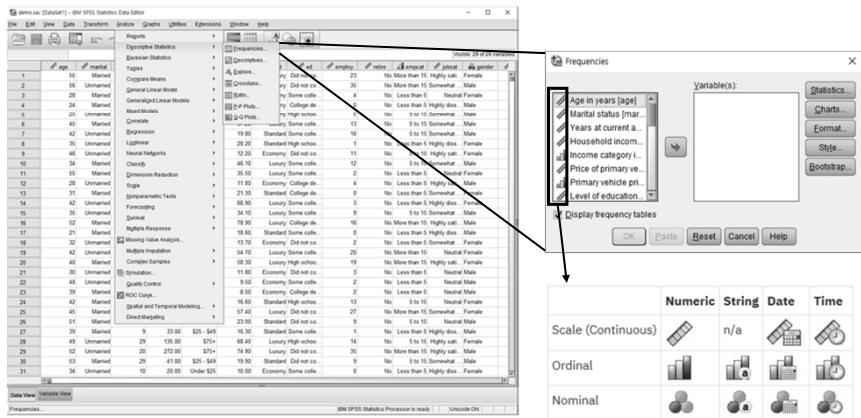
- Step 3: “DataSet” window allows you to edit the data variable using the variable View Tap



# IBM SPSS Statistics Project

## ❖ Project 1: IBM SPSS Statistics Subscription

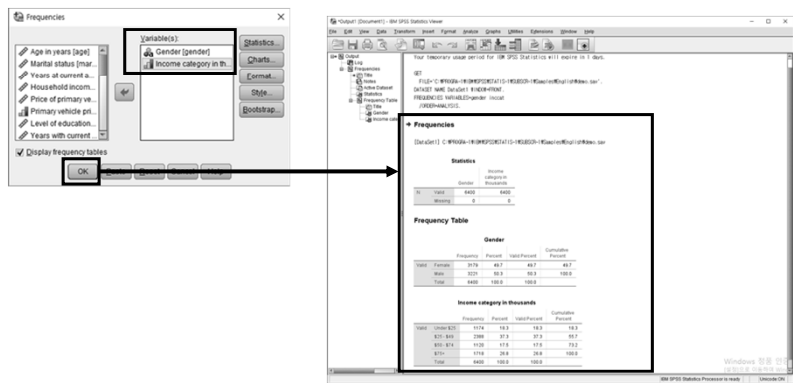
- Step 4-1: Run the Analysis



# IBM SPSS Statistics Project

## ❖ Project 1: IBM SPSS Statistics Subscription

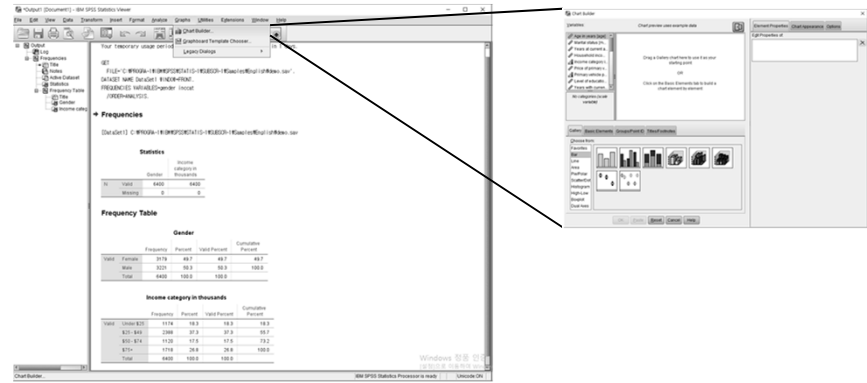
- Step 4-2: Chose variable Information and click “OK”



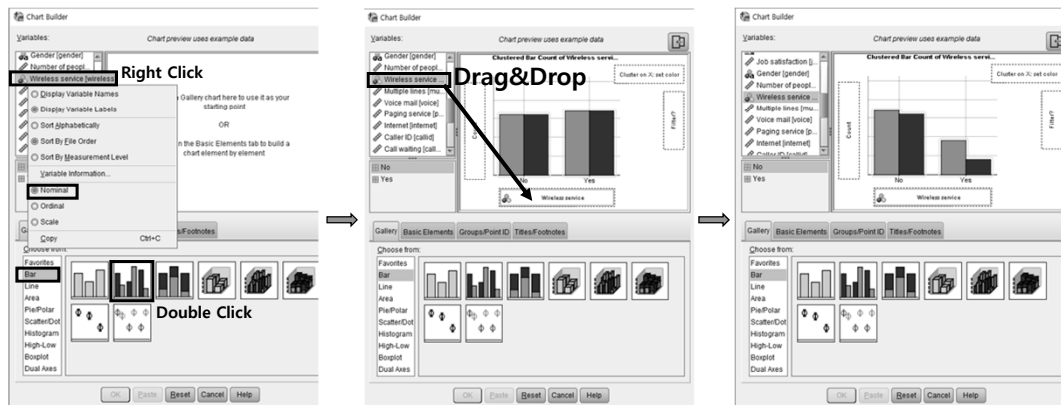
# IBM SPSS Statistics Project

## ❖ Project 1: IBM SPSS Statistics Subscription

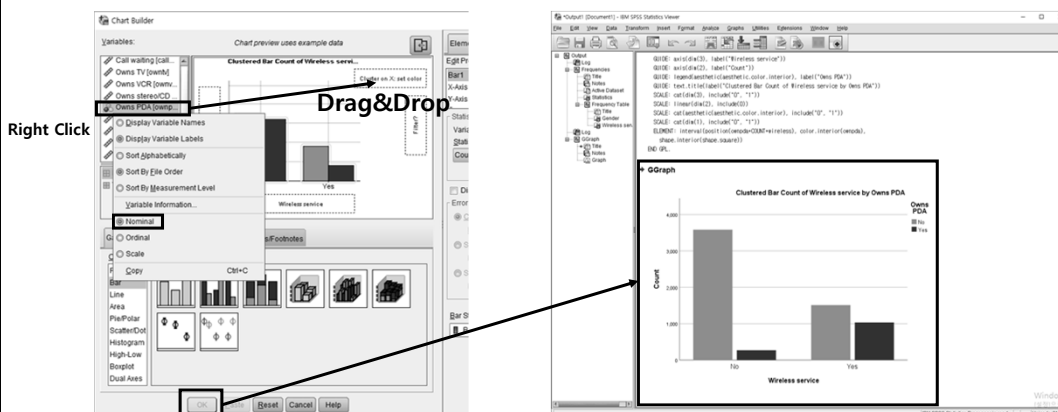
- Step 4-3: Graphs



- Step 4-3: Graphs



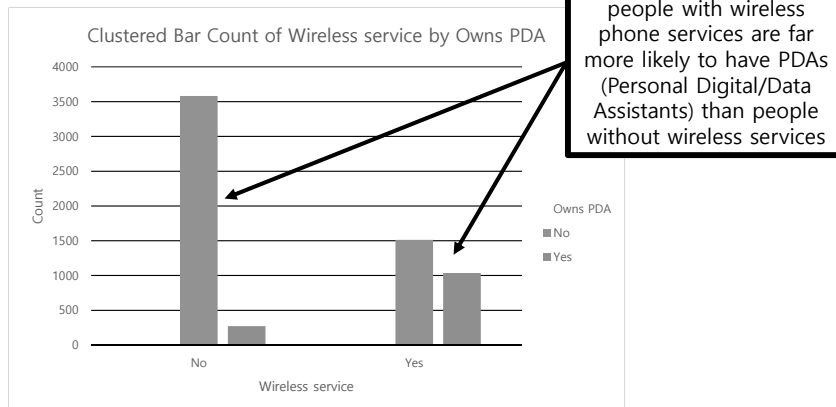
- Step 4-3: Graphs



## IBM SPSS Statistics Project

### ❖ Project 1: IBM SPSS Statistics Subscription

#### ▪ Step 5: Insight



## Project Requirements for Peer Review

### ❖ Project 2: Try to Analyze These

1. Analyze and Draw a Chart of the relationship between the 'Number of people in household' (reside) and the 'Primary vehicle price category' (carcat)
2. Analyze and Draw a Chart of the relationship between the 'Household income in thousands' (income) and the 'Primary vehicle price category' (carcat)
3. Compare the results of 1 and 2

## Project Requirements for Peer Review

- In order to succeed in the analysis of Project 2, you will need to compare the 'Number of people in household' and the 'Household income in thousands' based on the same 'Primary vehicle price category' reference data.
- Please answer this question
  - In reference to the Primary vehicle price category, is there correlation between Number of people and Household income?
- In addition, give helpful tips so your peers can learn more on IBM SPSS Statistics, as well as other corporate data analysis systems

## Big Data Emerging Technologies

### ❖ Modules

1. Big Data Rankings & Products
2. Big Data & Hadoop
3. Spark
4. Spark ML & Streaming
5. Storm
6. IBM SPSS Statistics Project

## Big Data References

### References

- <https://www.ibm.com/us-en/?lnk=m>
- <https://www.sap.com/corporate/en.html>
- <https://www.oracle.com/index.html>
- <https://www.hpe.com/us/en/home.html>
- <https://www.splunk.com/>
- <https://www.accenture.com/us-en/new-applied-now>
- <http://www.dell.com/en-us/?refid=df&s=gen&-ck=cr>
- <http://www.teradata.com/>
- <https://www.microsoft.com/en-us/>
- <https://www.cisco.com/>
- [https://aws.amazon.com/free/?nc1=h\\_ls](https://aws.amazon.com/free/?nc1=h_ls)
- ***Special Thanks to my Teaching Assistants***
  - *Jinbae Lee & Younghwan Shin*