Moving Beyond the Basics

Matthew Renze
SOFTWARE CONSULTANT

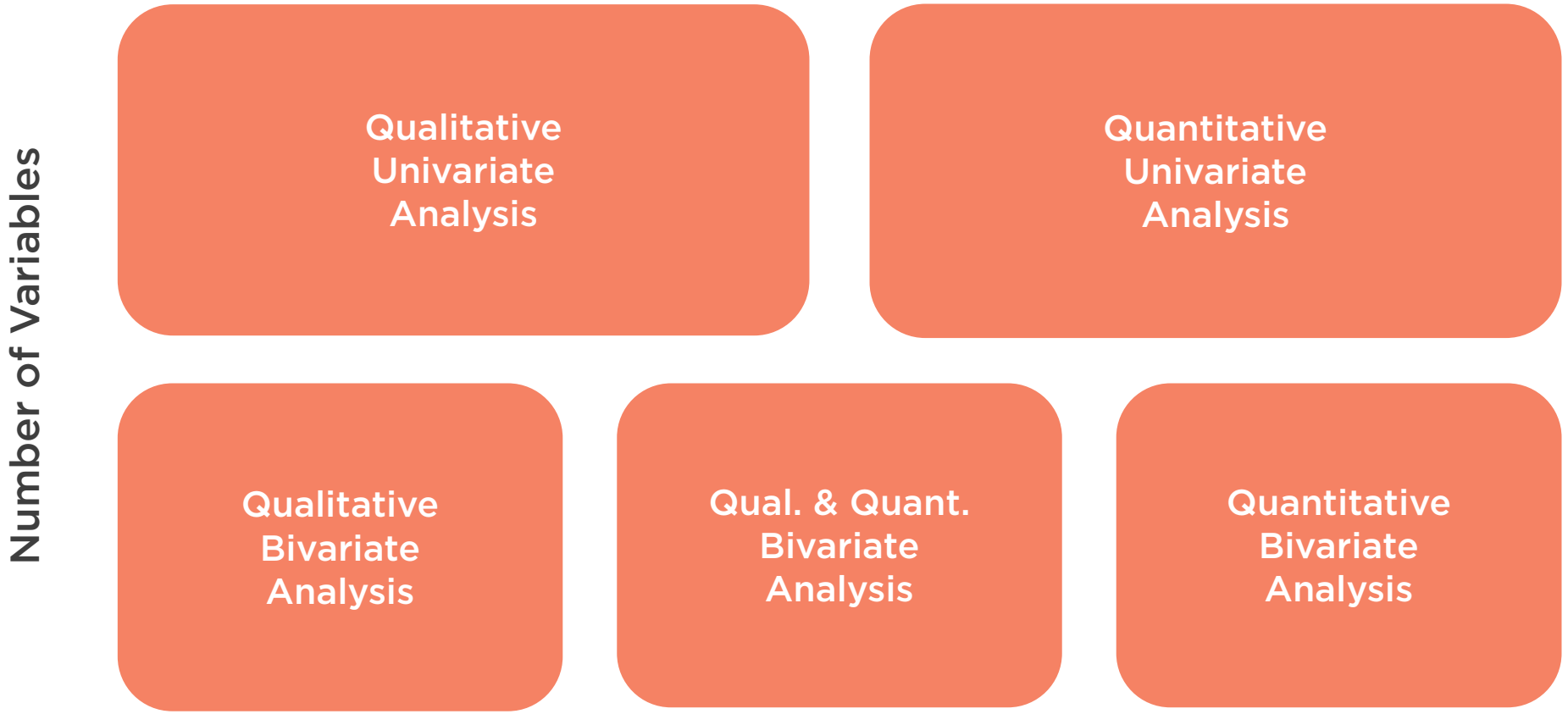
@matthewrenze www.matthewrenze.com

Overview

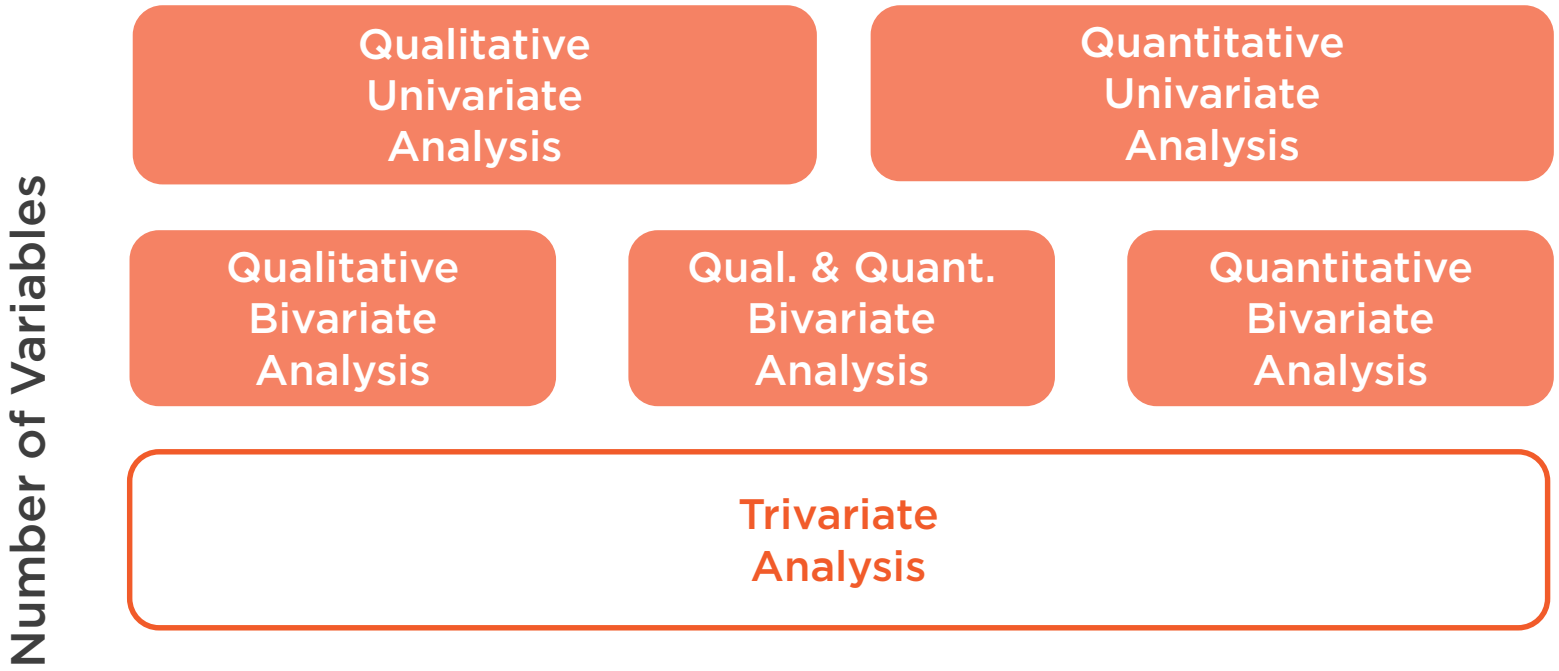
- Advanced Data Visualization
- Best Practices
- Demo
- Alternatives to R
- Course Summary


This is just the tip of the iceberg!

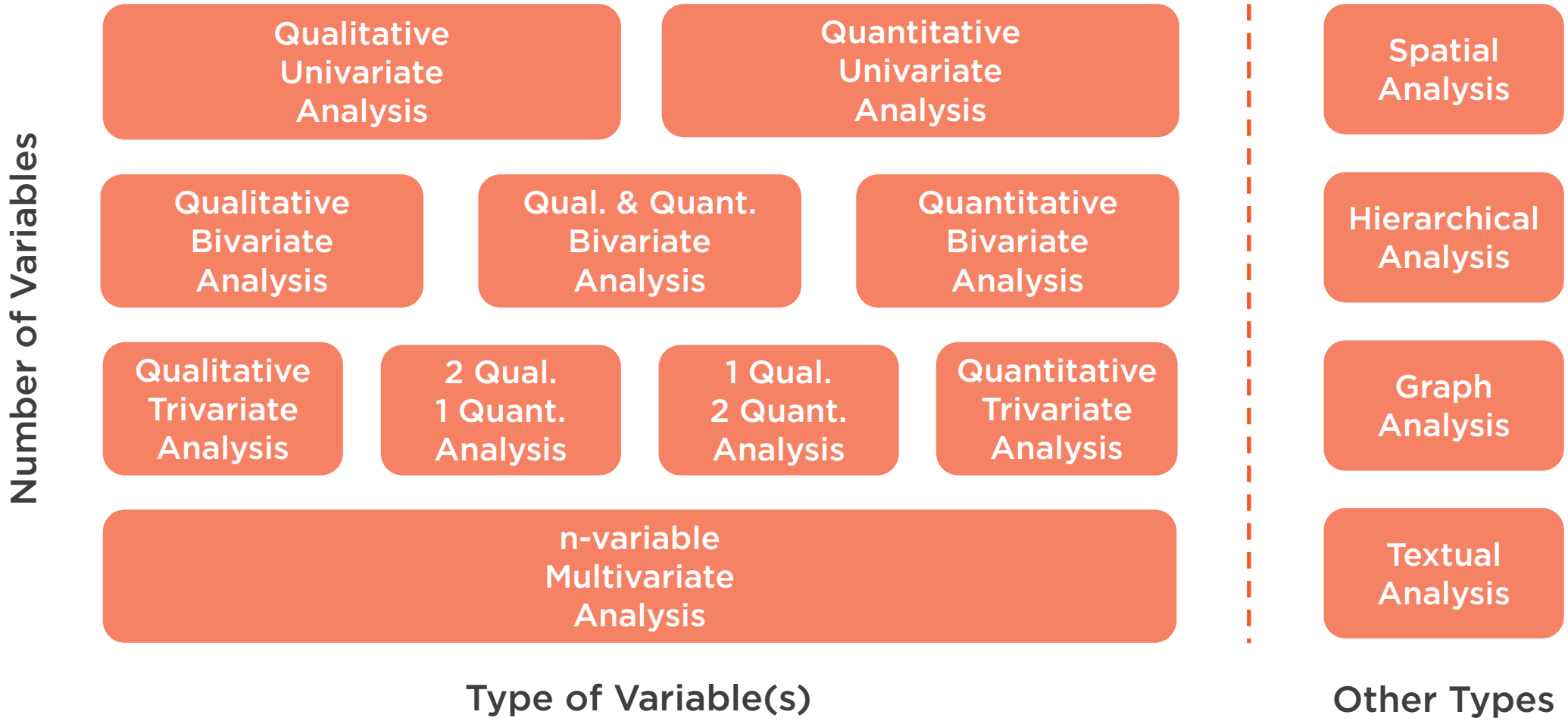
Types of Data Analysis



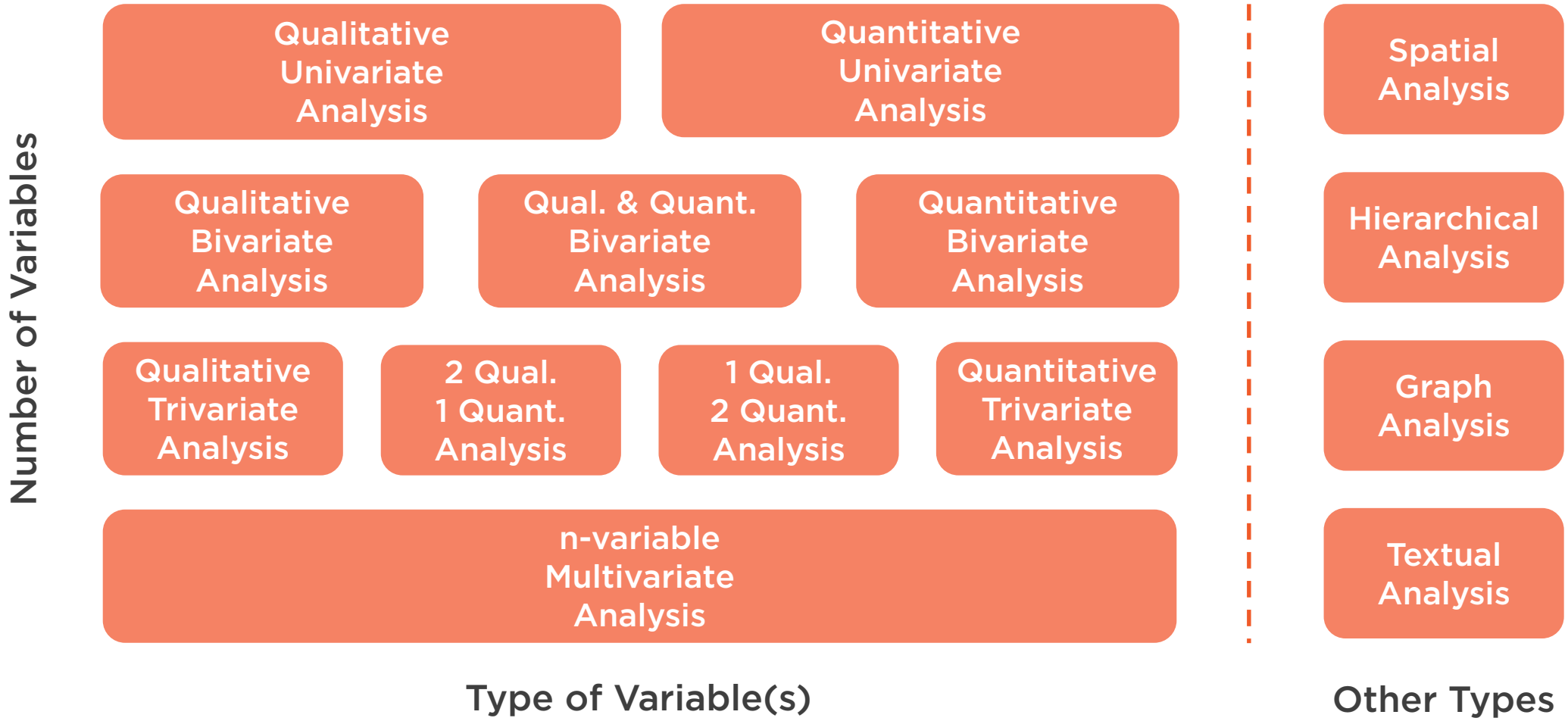
Types of Data Analysis



Types of Data Analysis



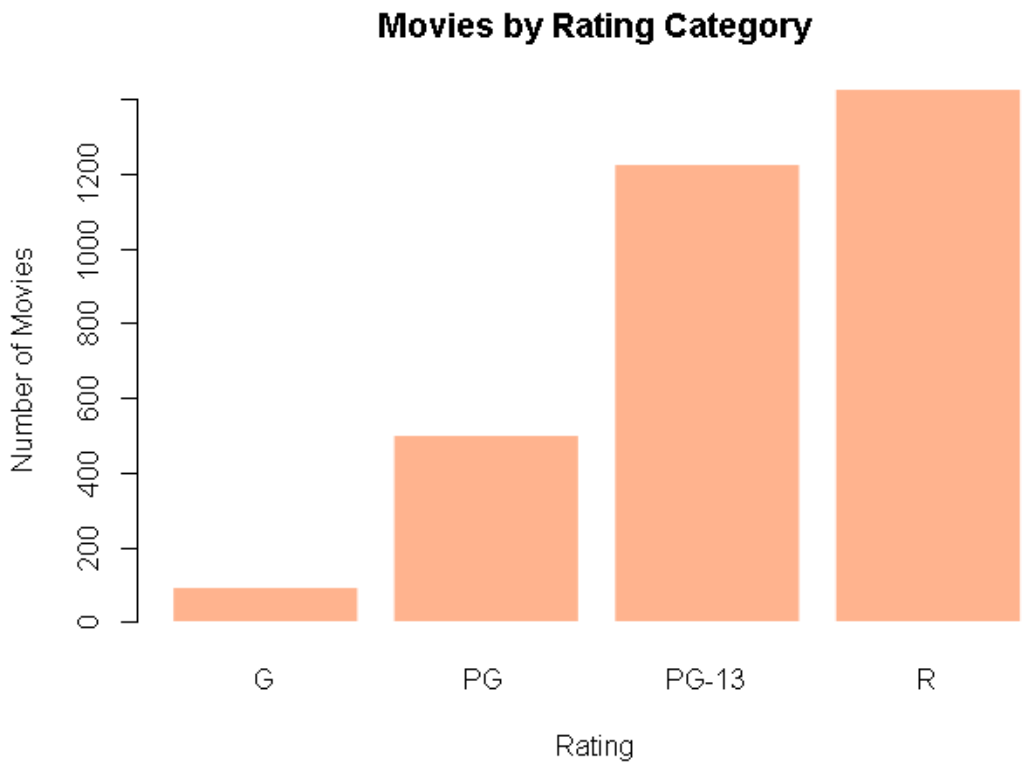
Types of Data Analysis



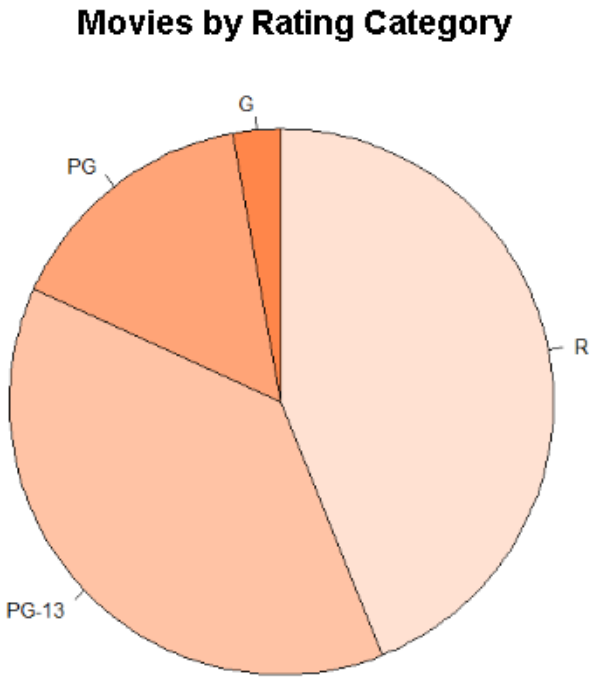

ADVICE
HELP
TIPS
SUPPORT

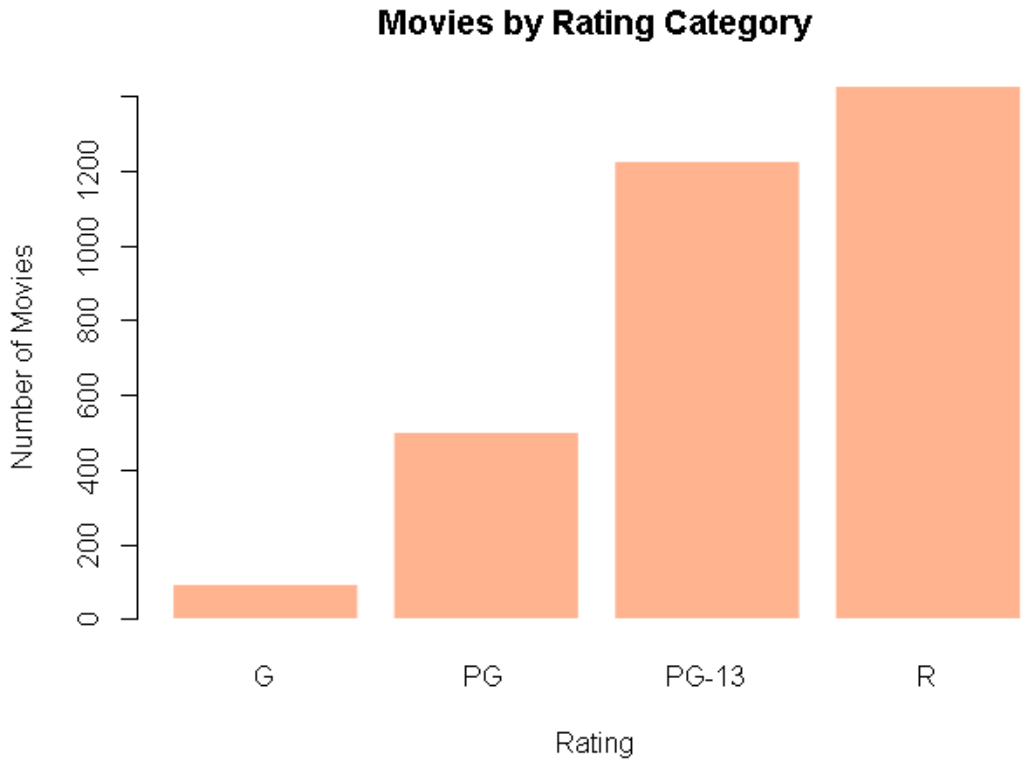
Start with a Question

How many movies were released in each rating category from 2000 to 2015?

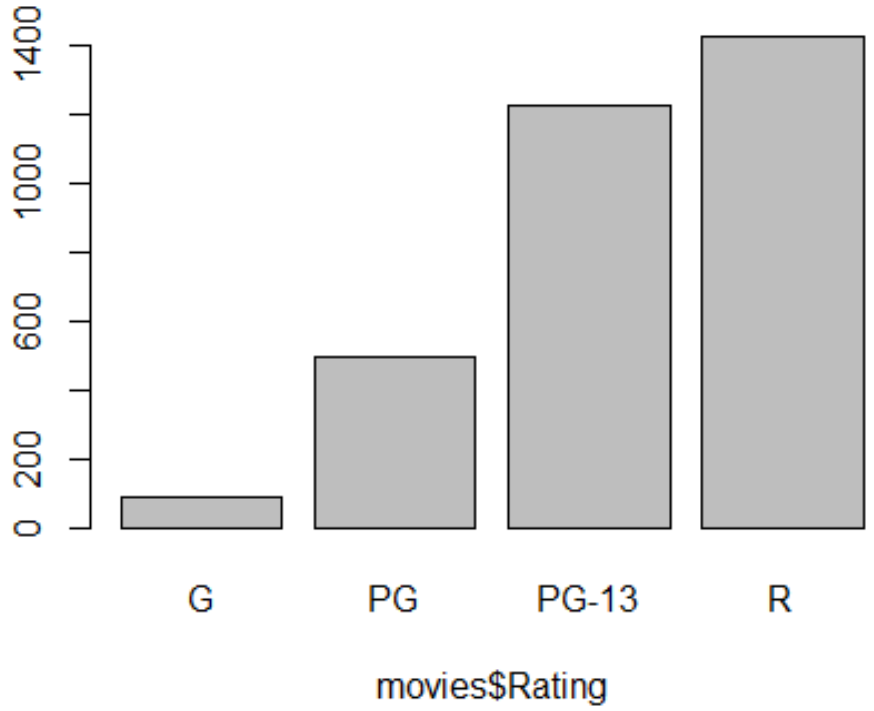


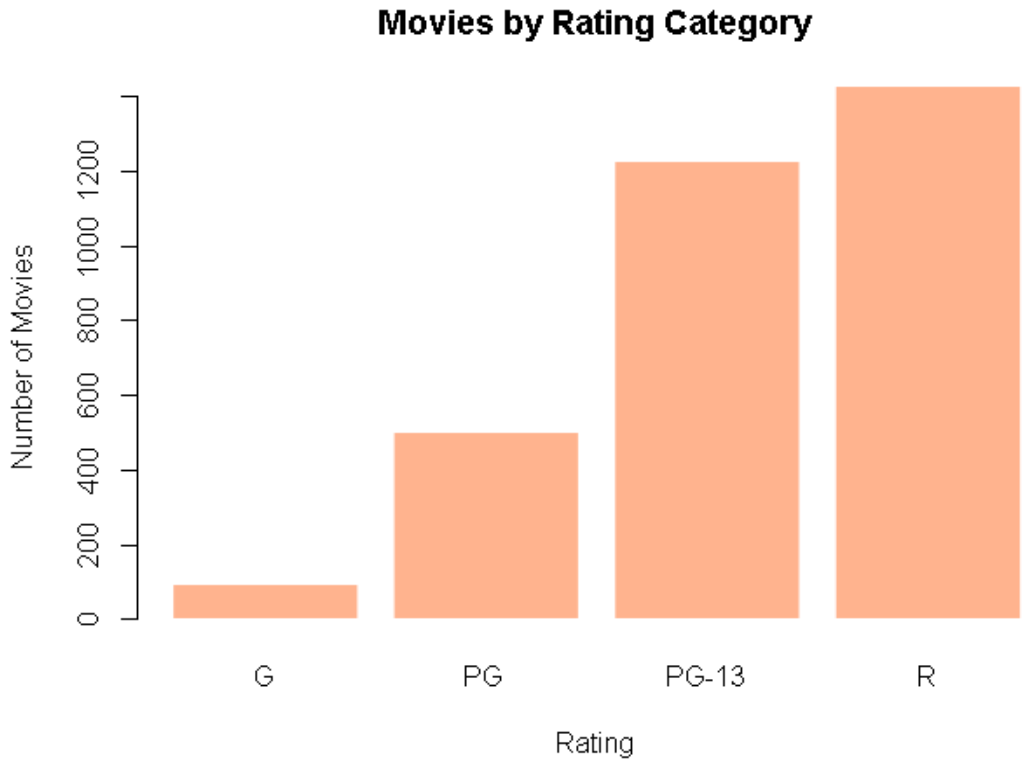
Use the Right Tool for the Job



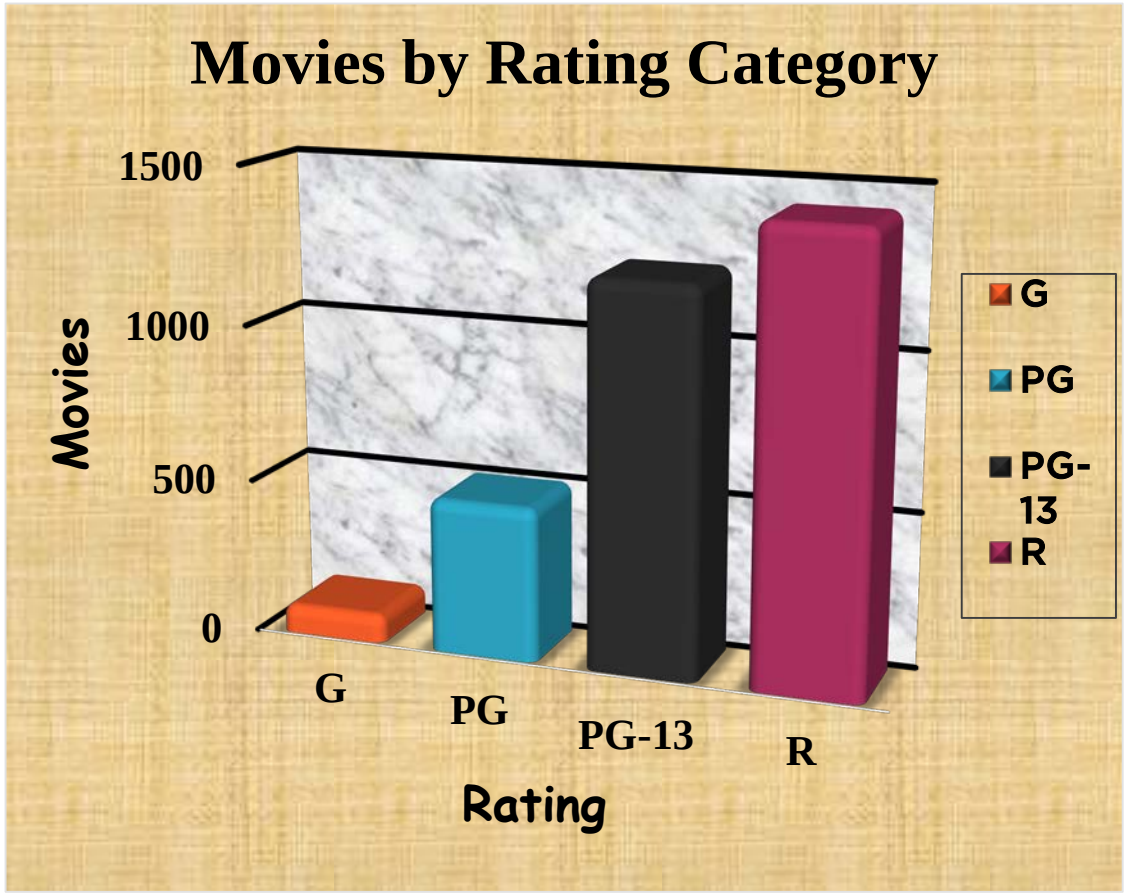


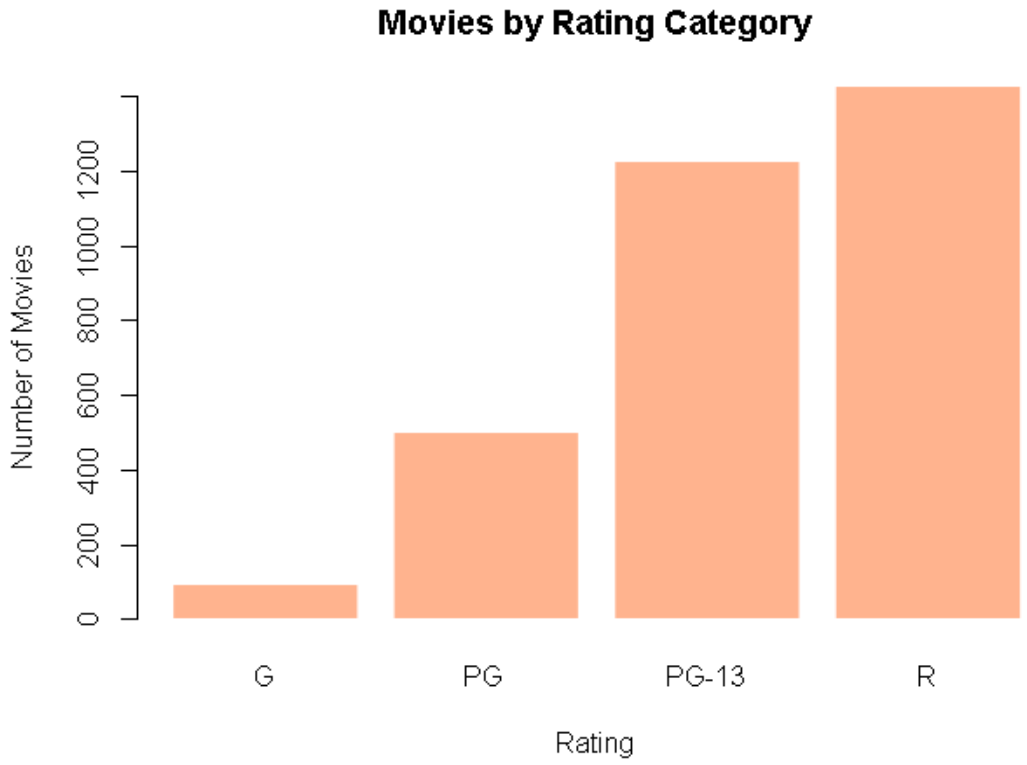
Know Your Audience



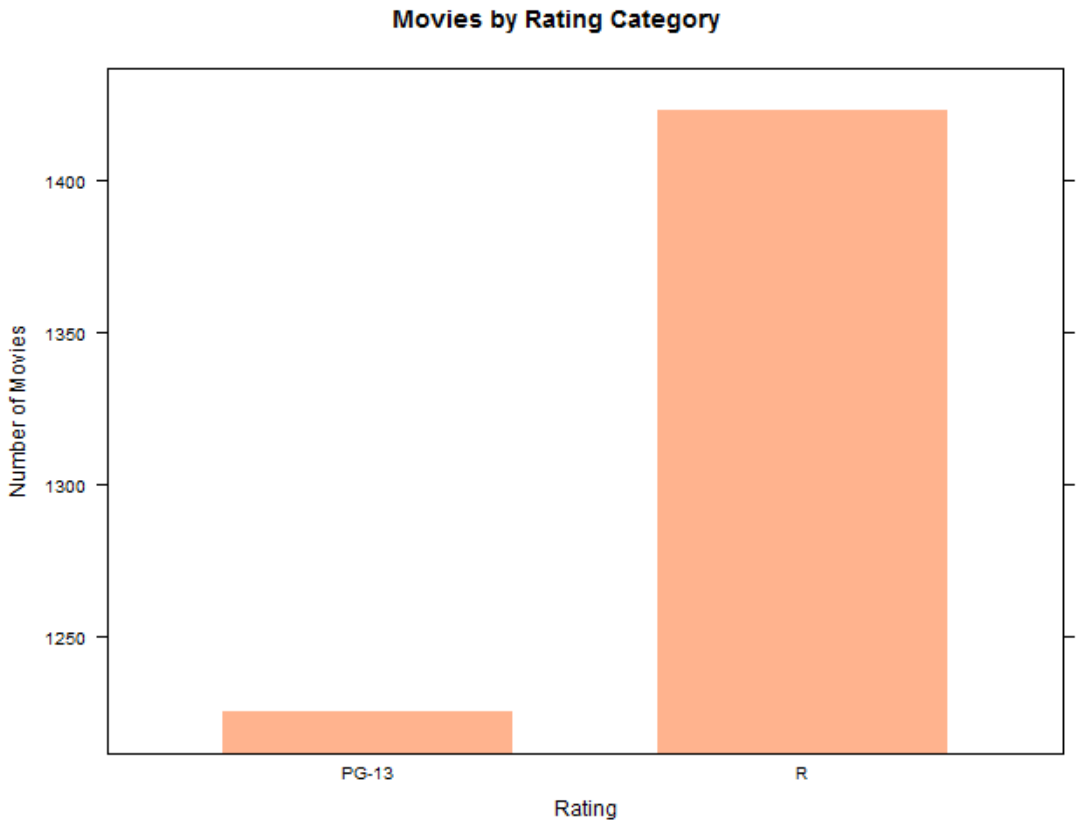


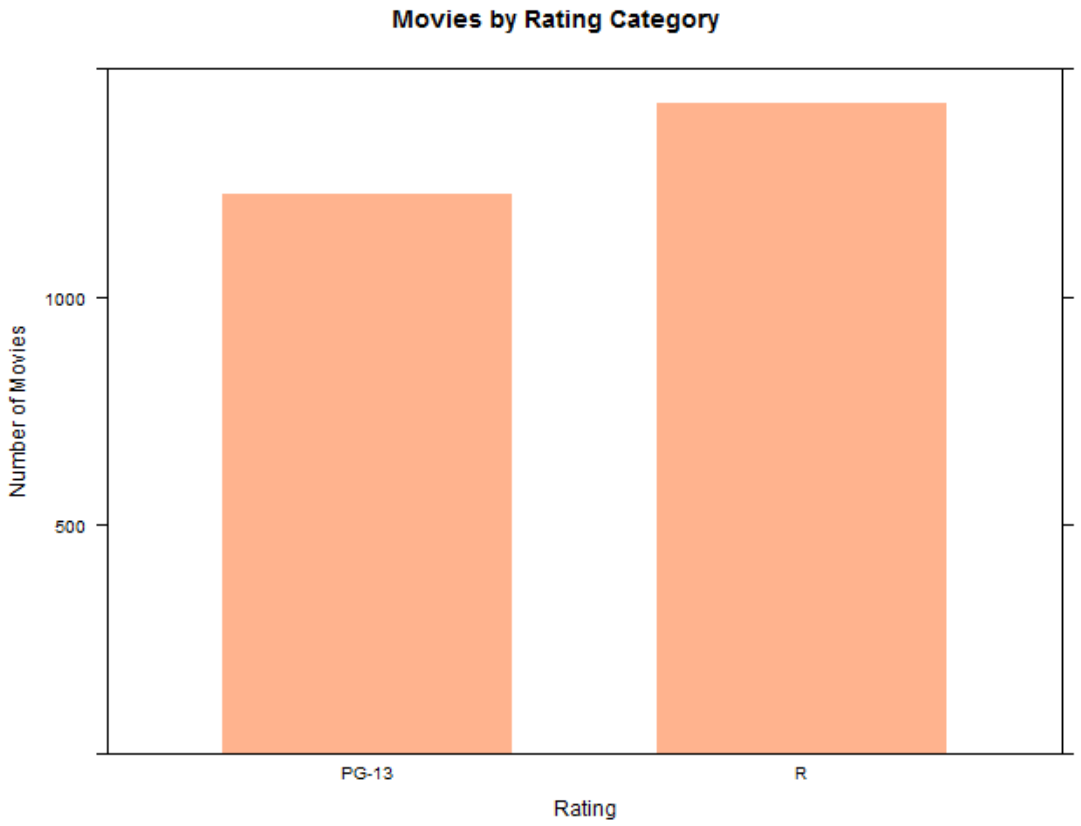
Create Clean Data Visualizations





Avoid Information Distortions





Create Bar Chart with Defaults

Add Context

Clean up the Data Visualization

Avoid Information Distortions

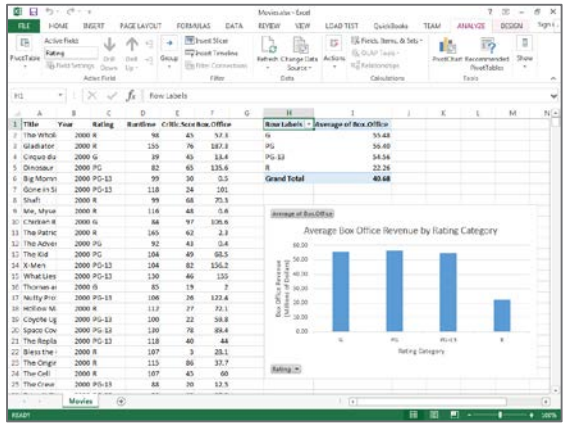
Export the Data Visualization

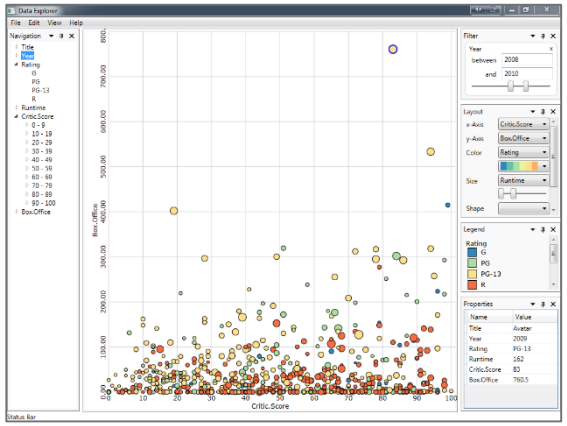
Export Using RStudio

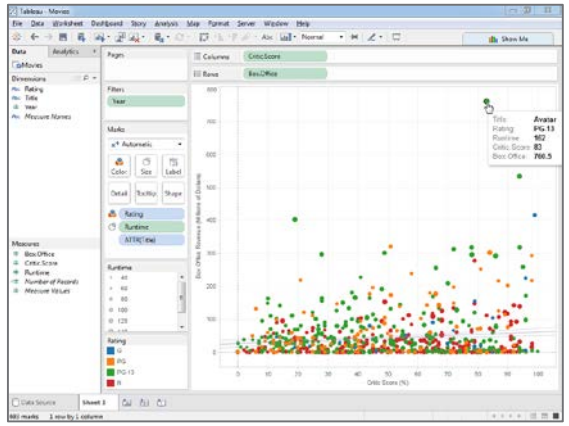
Export using ggsave

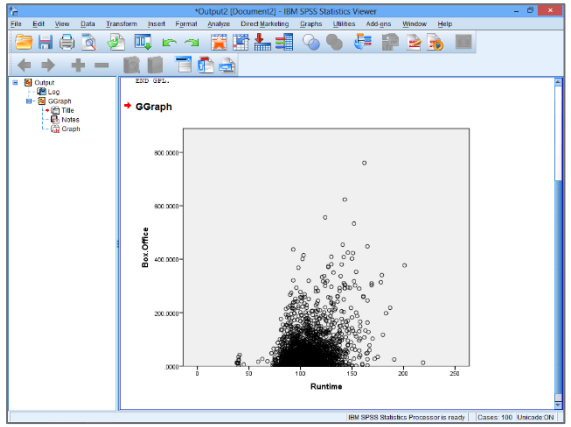
Viewing Help Files

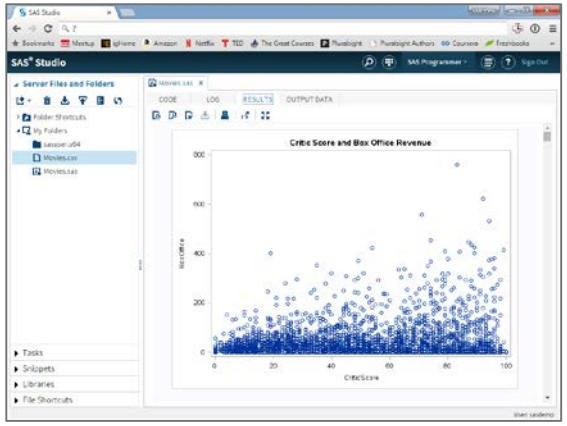
Alternatives to R for Data Visualization

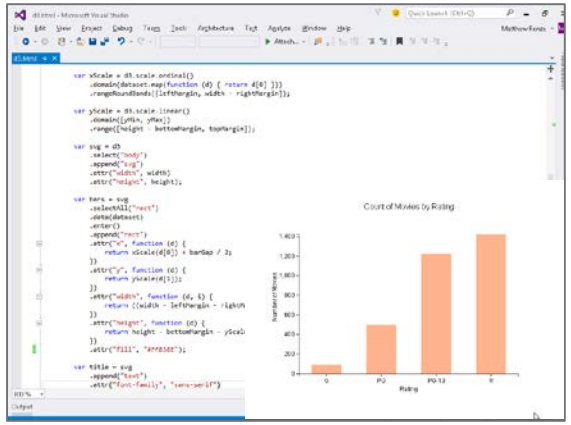


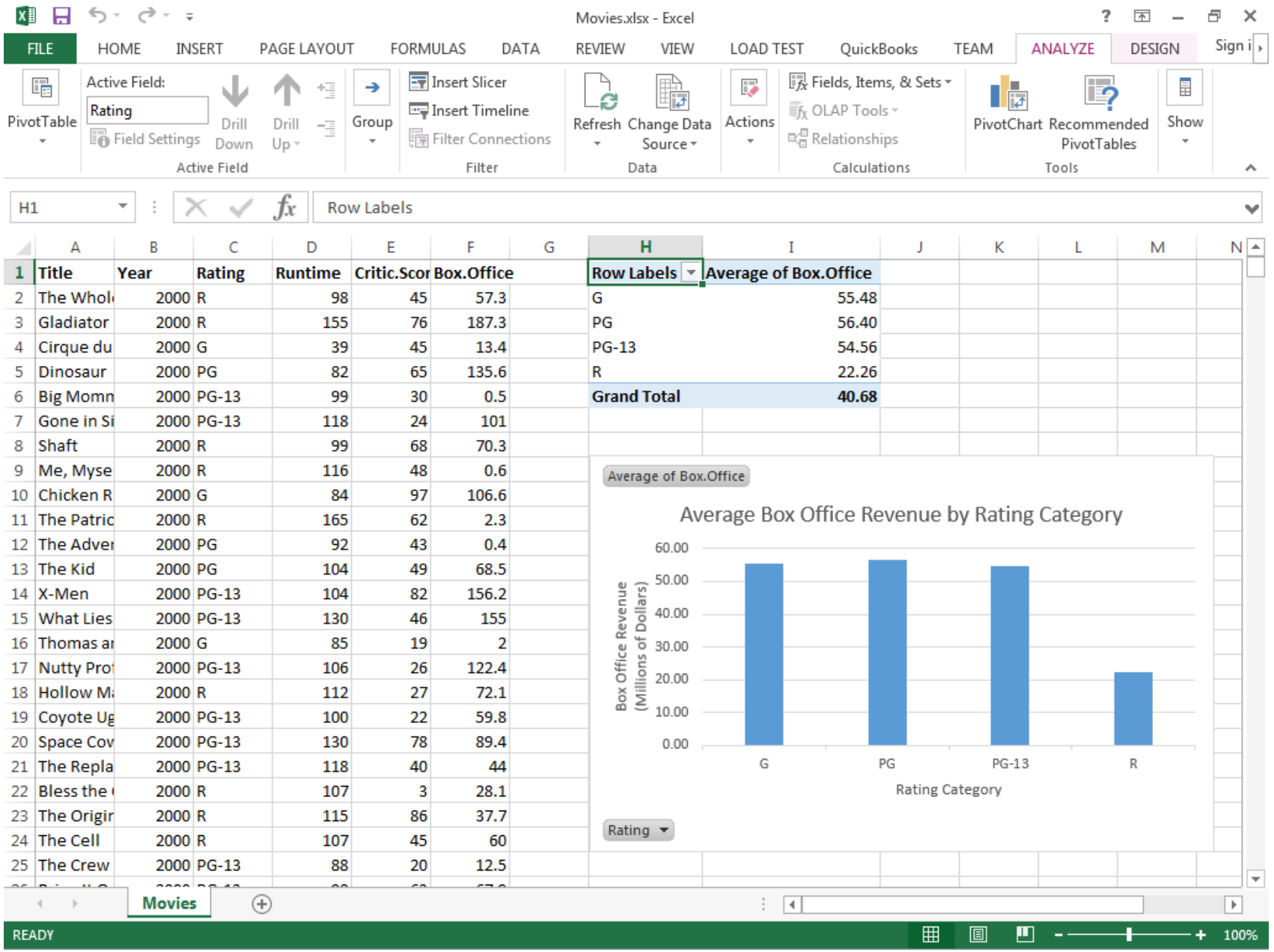




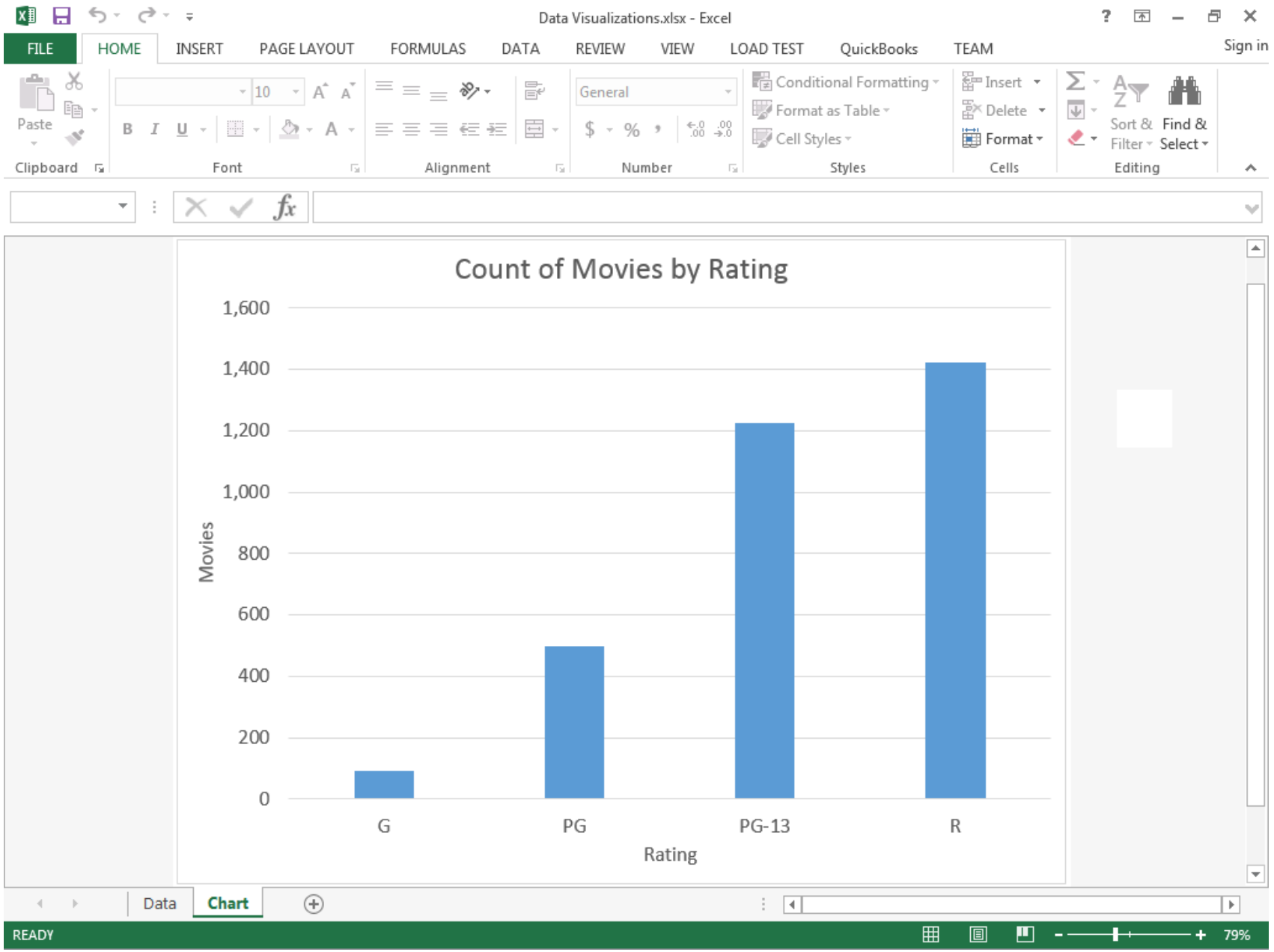


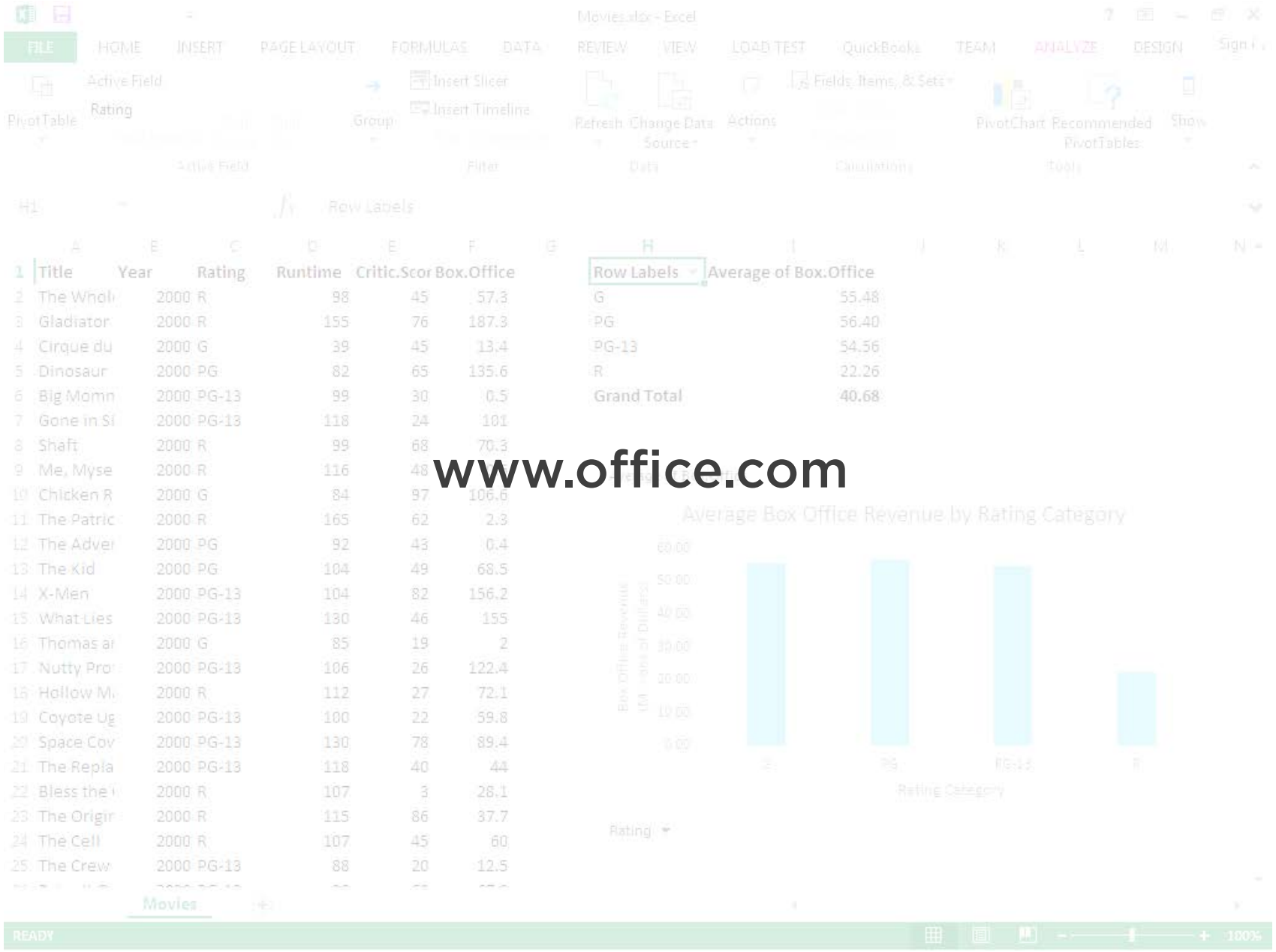


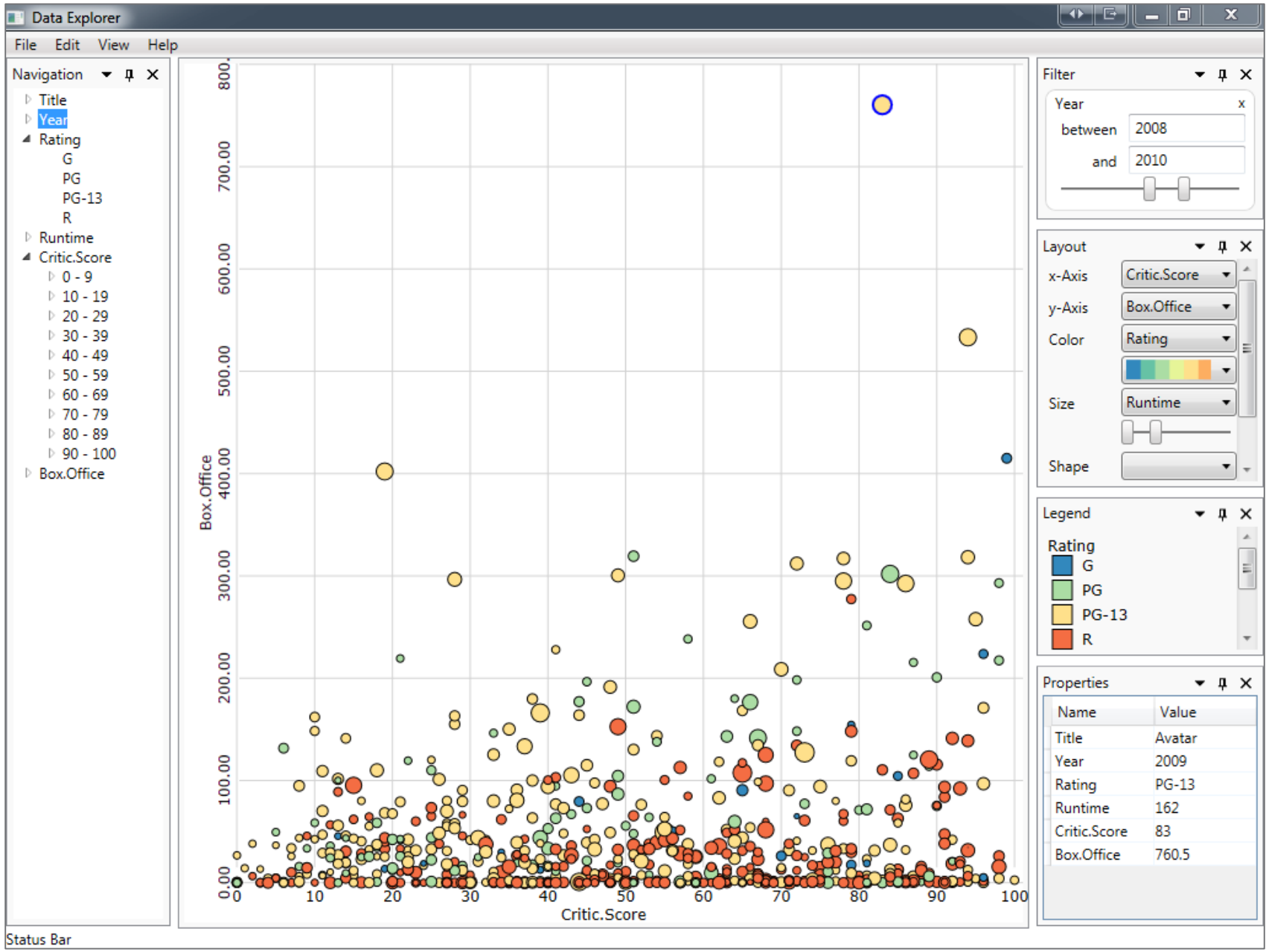


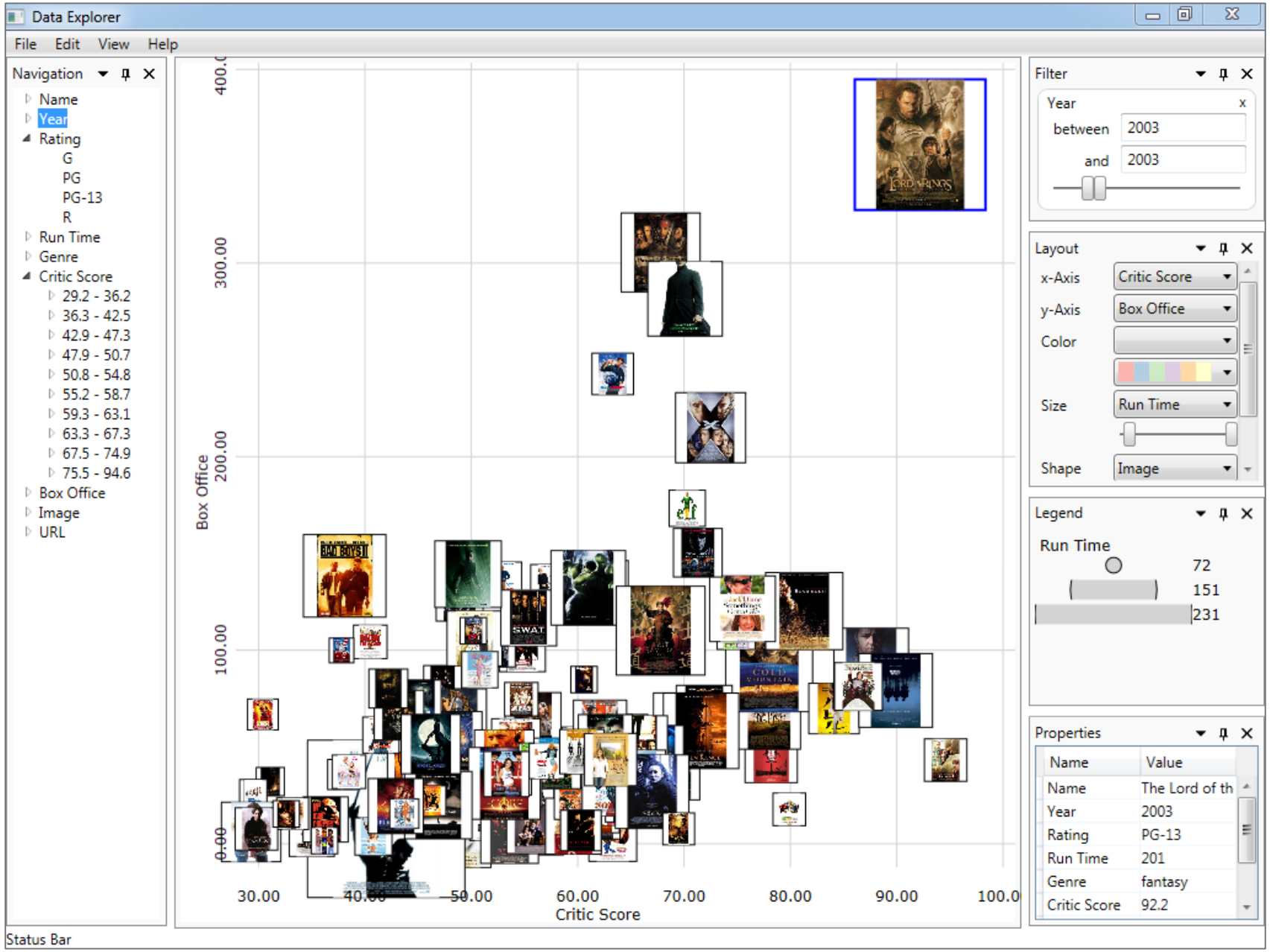


	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Title	Year	Rating	Runtime	Critic.Scor	Box.Office		Row Labels	Average of Box.Office					
2	The Whol	2000	R	98	45	57.3		G	55.48					
3	Gladiator	2000	R	155	76	187.3		PG	56.40					
4	Cirque du	2000	G	39	45	13.4		PG-13	54.56					
5	Dinosaur	2000	PG	82	65	135.6		R	22.26					
6	Big Momn	2000	PG-13	99	30	0.5		Grand Total	40.68					
7	Gone in Si	2000	PG-13	118	24	101								
8	Shaft	2000	R	99	68	70.3								
9	Me, Myse	2000	R	116	48	0.6								
10	Chicken R	2000	G	84	97	106.6								
11	The Patric	2000	R	165	62	2.3								
12	The Adver	2000	PG	92	43	0.4								
13	The Kid	2000	PG	104	49	68.5								
14	X-Men	2000	PG-13	104	82	156.2								
15	What Lies	2000	PG-13	130	46	155								
16	Thomas ar	2000	G	85	19	2								
17	Nutty Prof	2000	PG-13	106	26	122.4								
18	Hollow Ma	2000	R	112	27	72.1								
19	Coyote Ug	2000	PG-13	100	22	59.8								
20	Space Cov	2000	PG-13	130	78	89.4								
21	The Repla	2000	PG-13	118	40	44								
22	Bless the	2000	R	107	3	28.1								
23	The Origir	2000	R	115	86	37.7								
24	The Cell	2000	R	107	45	60								
25	The Crew	2000	PG-13	88	20	12.5								


Data Visualizations.xlsx - Excel
FILE
HOME
INSERT
PAGE LAYOUT
FORMULAS
DATA
REVIEW
VIEW
LOAD TEST
QuickBooks
TEAM
Sign in
Paste
Clipboard
Font
Alignment
Number
Conditional Formatting
Format as Table
Cell Styles
Styles
Insert
Delete
Format
Cells
Sort & Filter
Find & Select
Editing
Count of Movies by Rating
1,600
1,400
1,200
1,000
800
600
400
200
0
Movies
G
PG
PG-13
R
Rating
Data
Chart
READY
79%


Movies.xlsx - Excel
FILE
HOME
INSERT
PAGE LAYOUT
FORMULAS
DATA
REVIEW
VIEW
LOAD TEST
QuickBooks
TEAM
ANALYZE
DESIGN
Active Field
Rating
PivotTable
Group
Insert Slicer
Insert Timeline
Refresh
Change Data Source
Actions
Fields, Items, & Sets
PivotChart
Recommended PivotTables
Show
Active Field
Filter
Data
Calculations
Tools
H1
Row Labels
Title
Year
Rating
Runtime
Critic.Scor
Box.Office
Row Labels
Average of Box.Office
G
55.48
PG
56.40
PG-13
54.56
R
22.26
Grand Total
40.68
www.office.com
Average Box Office Revenue by Rating Category
Box Office Revenue (in Tens of Dollars)
Rating Category
Rating
Movies
READY
100%


Data Explorer
File
Edit
View
Help
Navigation
Title
Year
Rating
G
PG
PG-13
R
Runtime
Critic.Score
0 - 9
10 - 19
20 - 29
30 - 39
40 - 49
50 - 59
60 - 69
70 - 79
80 - 89
90 - 100
Box.Office
Box.Office
0.00
100.00
200.00
300.00
400.00
500.00
600.00
700.00
800.00
0
10
20
30
40
50
60
70
80
90
100
Critic.Score
Filter
Year
between
2008
and
2010
Layout
x-Axis
Critic.Score
y-Axis
Box.Office
Color
Rating
Size
Runtime
Shape
Legend
Rating
G
PG
PG-13
R
Properties
Name
Value
Title
Avatar
Year
2009
Rating
PG-13
Runtime
162
Critic.Score
83
Box.Office
760.5
Status Bar


Data Explorer
File
Edit
View
Help
Navigation
Name
Year
Rating
G
PG
PG-13
R
Run Time
Genre
Critic Score
29.2 - 36.2
36.3 - 42.5
42.9 - 47.3
47.9 - 50.7
50.8 - 54.8
55.2 - 58.7
59.3 - 63.1
63.3 - 67.3
67.5 - 74.9
75.5 - 94.6
Box Office
Image
URL
Box Office
400.00
300.00
200.00
100.00
0.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.0
Critic Score
Filter
Year
between
2003
and
2003
Layout
x-Axis
Critic Score
y-Axis
Box Office
Color
Size
Run Time
Shape
Image
Legend
Run Time
72
151
231
Properties
Name
Value
Name
The Lord of th
Year
2003
Rating
PG-13
Run Time
201
Genre
fantasy
Critic Score
92.2
Status Bar


Data Explorer
File Edit View Help
Navigation
Title
Rating
G
PG
PG-13
R
Runtime
Critic.Score
0 - 9
10 - 19
20 - 29
30 - 39
40 - 49
50 - 59
60 - 69
70 - 79
80 - 89
90 - 100
Box.Office
www.data-explorer.com
Box.Office
Critic.Score
Filter
Year
between 2008
and 2010
Layout
x-Axis Critic.Score
y-Axis Box.Office
Color Rating
Size Runtime
Shape
Legend
Rating
G
PG
PG-13
R
Properties
Name Value
Title Avatar
Year 2009
Rating PG-13
Runtime 162
Critic.Score 83
Box.Office 760.5
Status Bar

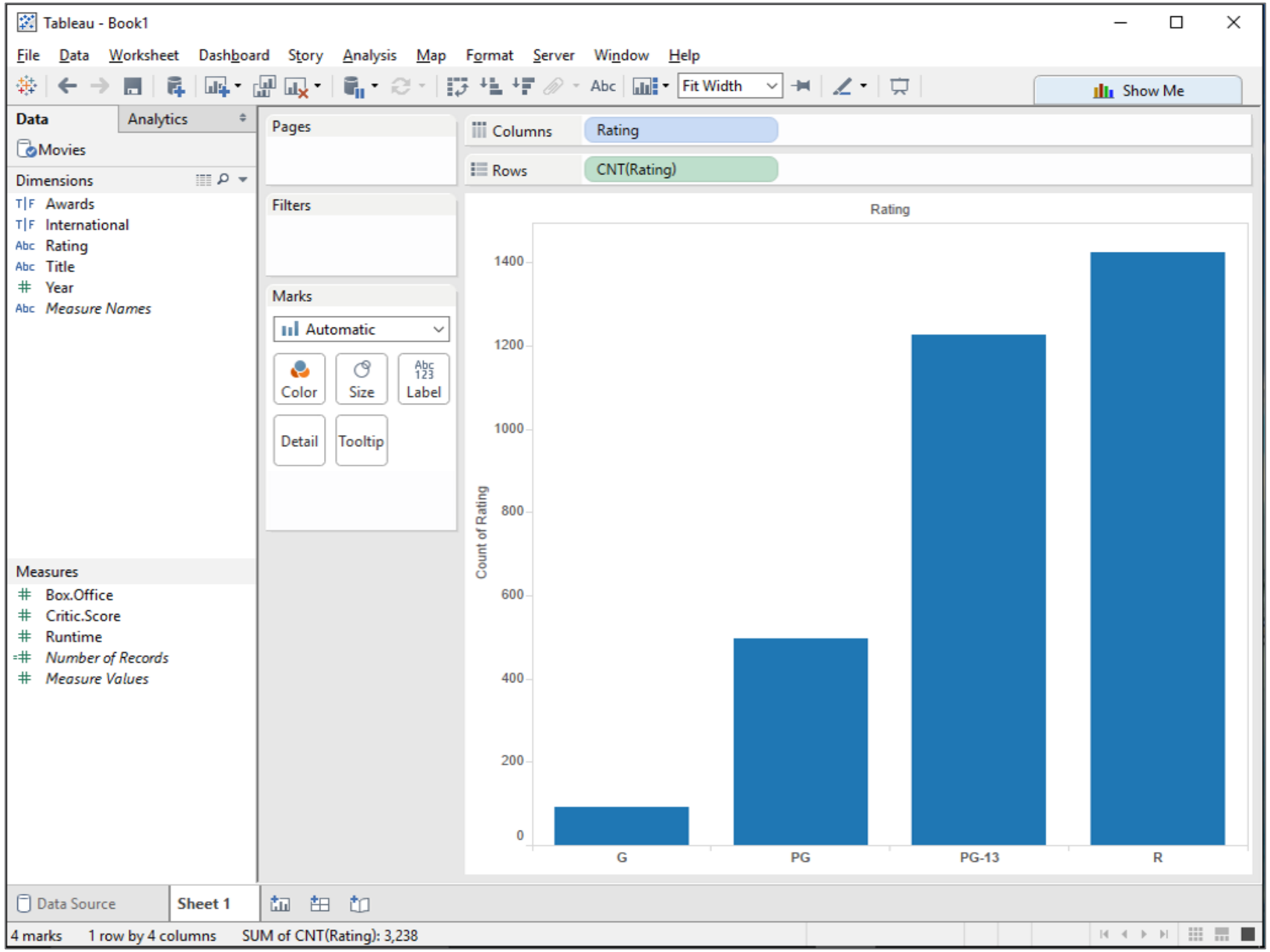


Tableau - Book1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Fit Width
Show Me
Data
Analytics
Movies
Dimensions
Awards
International
Rating
Title
Year
Measure Names
Measures
Box.Office
Critic.Score
Runtime
Number of Records
Measure Values
Pages
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Columns
Rating
Rows
CNT(Rating)
Rating
Count of Rating
1400
1200
1000
800
600
400
200
0
G
PG
PG-13
R
Data Source
Sheet 1
4 marks
1 row by 4 columns
SUM of CNT(Rating): 3,238

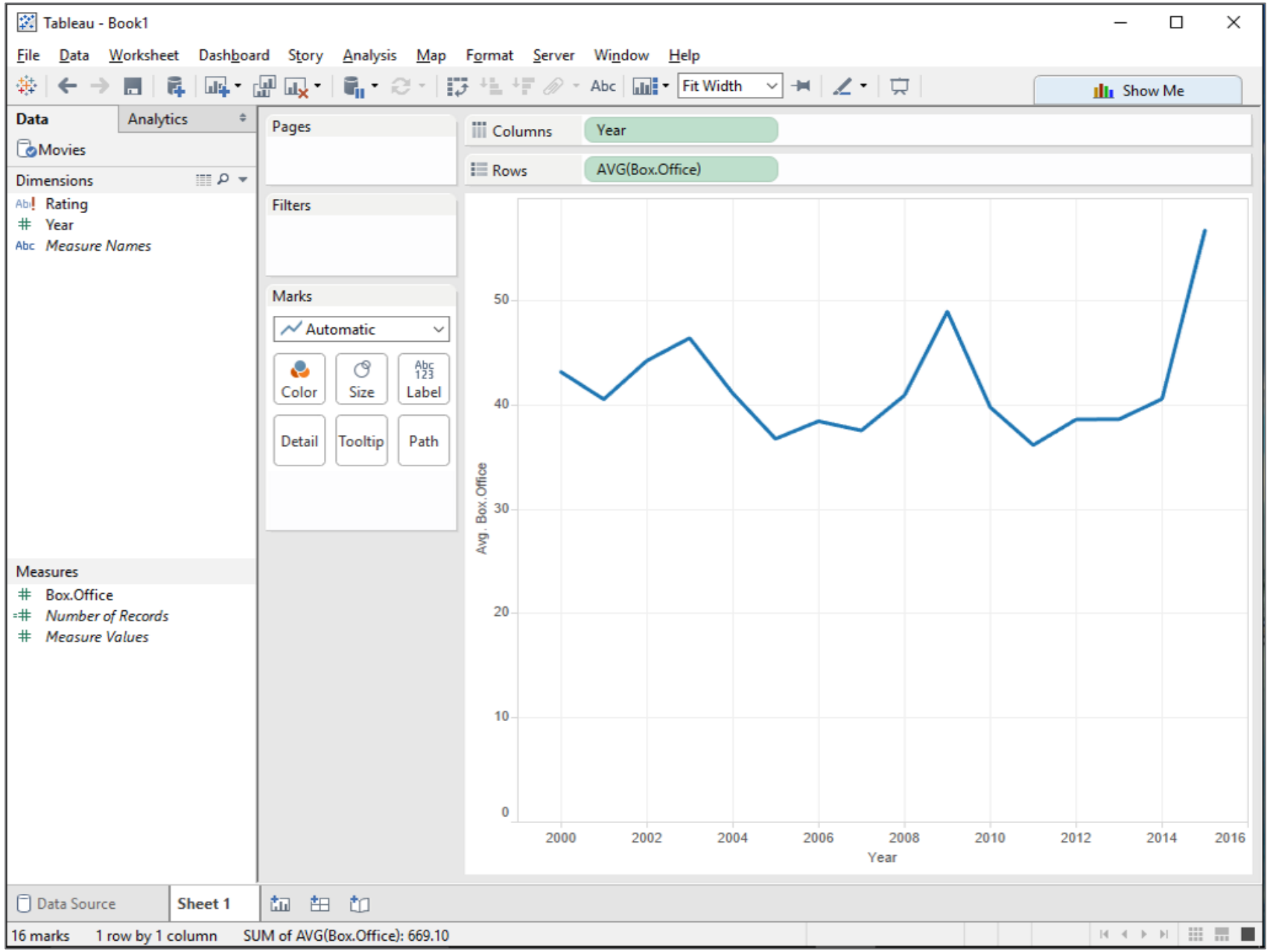


Tableau - Book1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Fit Width
Show Me
Data
Analytics
Movies
Dimensions
Rating
Year
Measure Names
Measures
Box.Office
Number of Records
Measure Values
Pages
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Path
Columns
Year
Rows
AVG(Box.Office)
Avg. Box.Office
0
10
20
30
40
50
2000
2002
2004
2006
2008
2010
2012
2014
2016
Year
Data Source
Sheet 1
16 marks
1 row by 1 column
SUM of AVG(Box.Office): 669.10

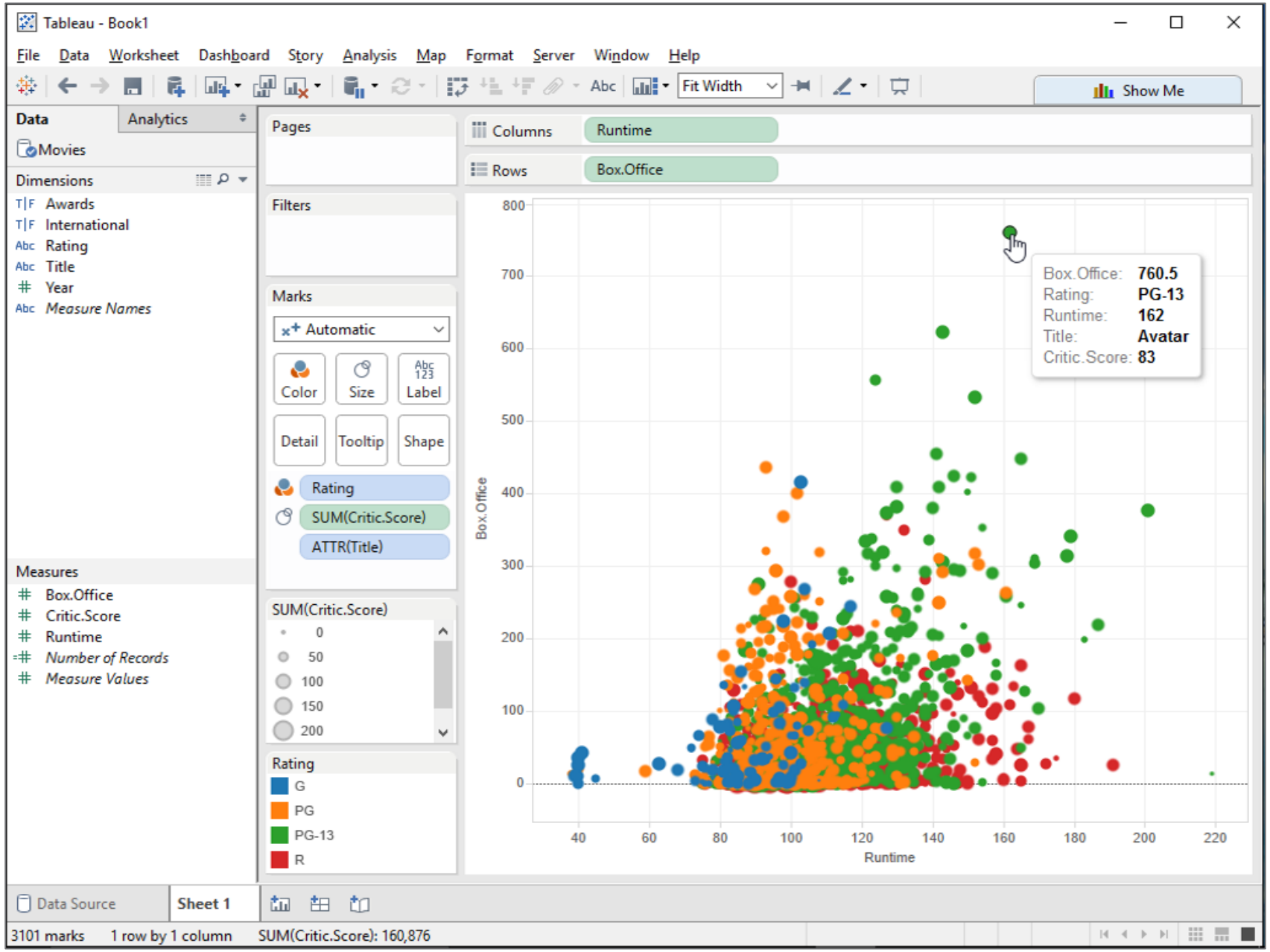
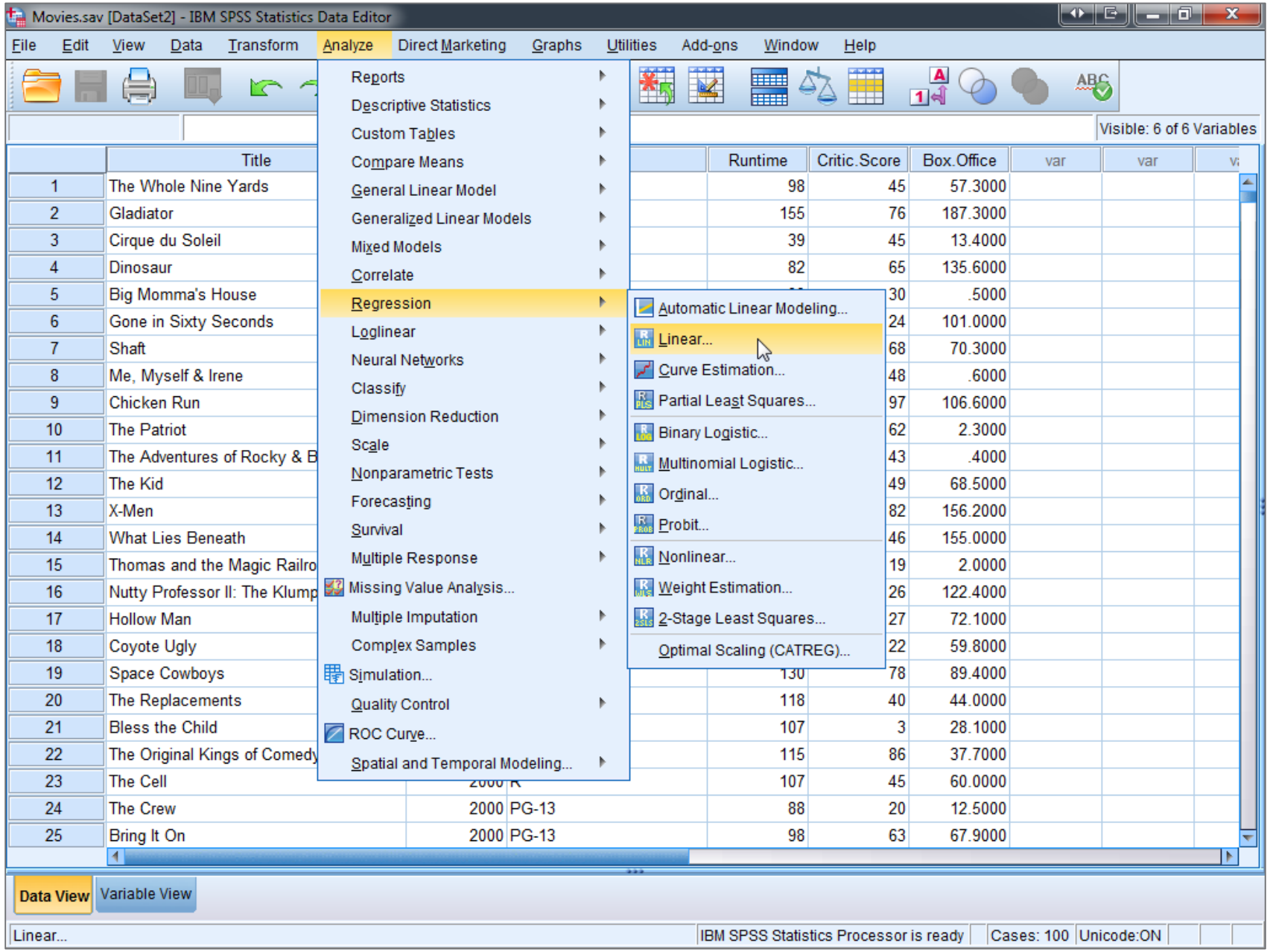
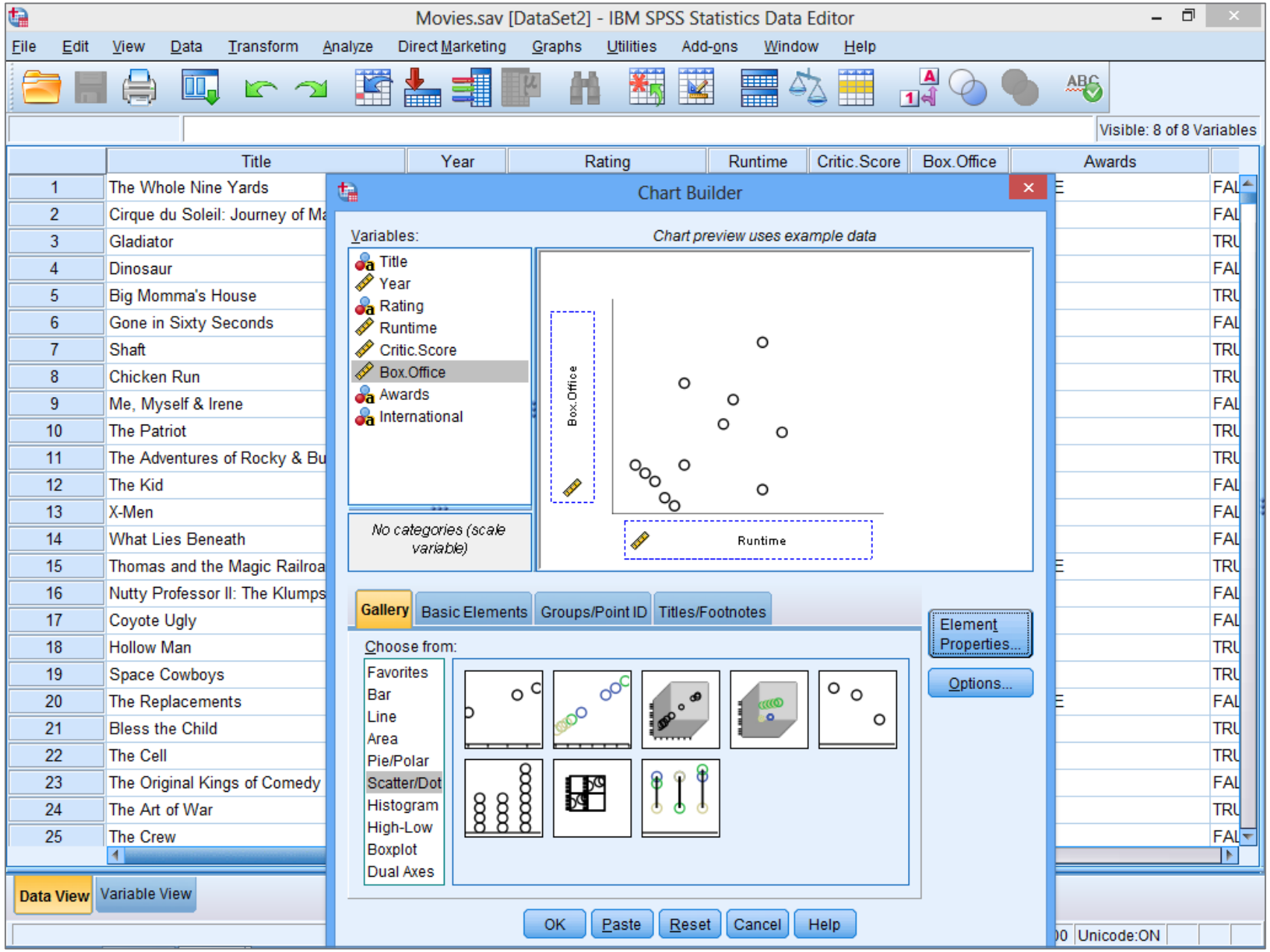
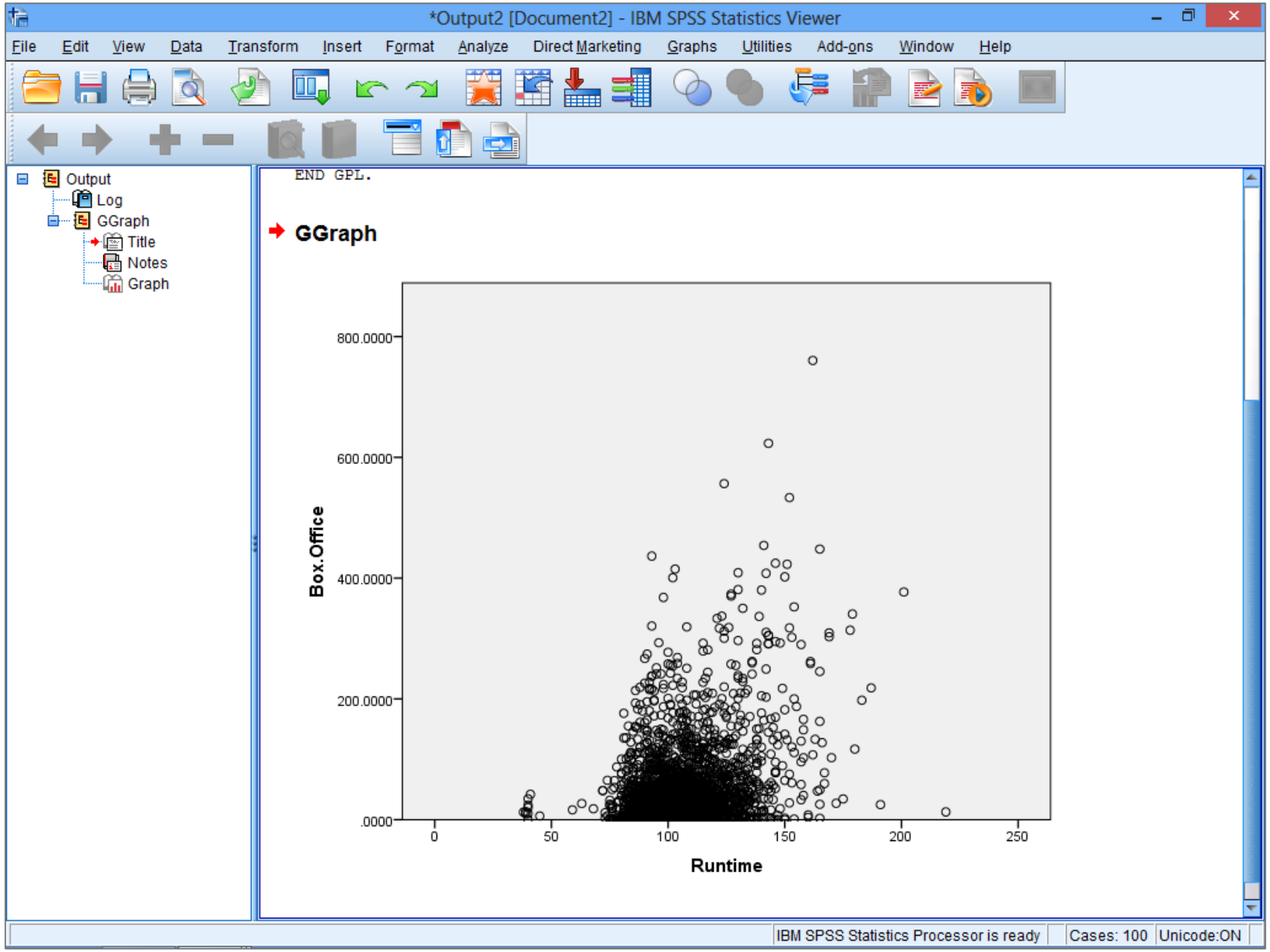


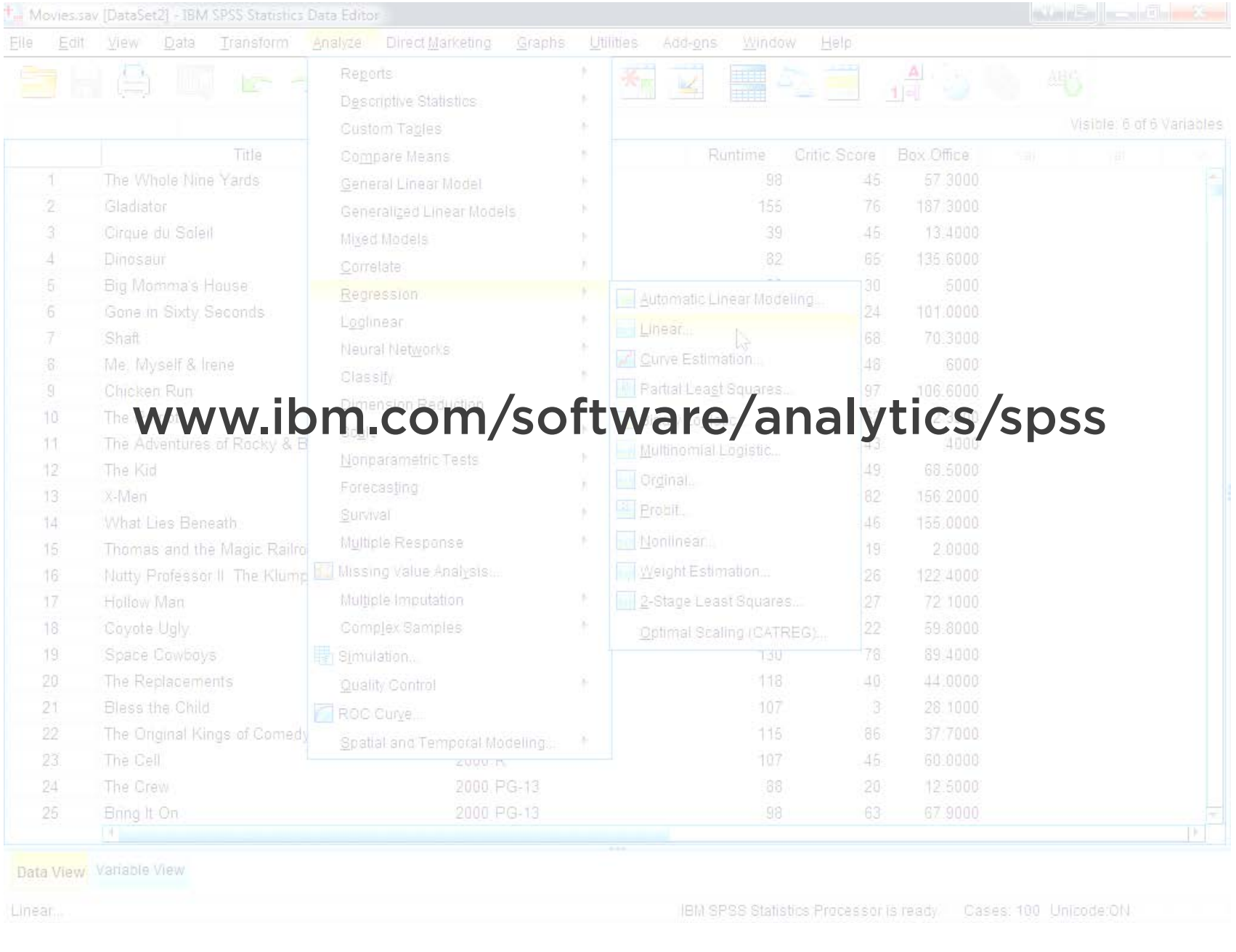
Tableau - Book1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Fit Width
Show Me
Data
Analytics
Movies
Dimensions
Awards
International
Rating
Title
Year
Measure Names
Measures
Box.Office
Critic.Score
Runtime
Number of Records
Measure Values
Pages
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Shape
Rating
SUM(Critic.Score)
ATTR(Title)
SUM(Critic.Score)
0
50
100
150
200
Rating
G
PG
PG-13
R
Columns
Runtime
Rows
Box.Office
Box.Office: 760.5
Rating: PG-13
Runtime: 162
Title: Avatar
Critic.Score: 83
800
700
600
500
400
300
200
100
0
Box.Office
40
60
80
100
120
140
160
180
200
220
Runtime
Data Source
Sheet 1
3101 marks
1 row by 1 column
SUM(Critic.Score): 160,876


Tableau - Movies
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Normal
Show Me
Data
Analytics
Movies
Dimensions
Rating
Title
Year
Measure Names
Pages
Columns
Critic.Score
Rows
Box.Office
Filters
Year
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Shape
Rating
Runtime
ATTR(Title)
Measures
Box.Office
Critic.Score
Runtime
Number of Records
Measure Values
Avatar
PG-13
162
83
760.5
www.tableau.com
Data Source
Sheet 1
693 marks
1 row by 1 column

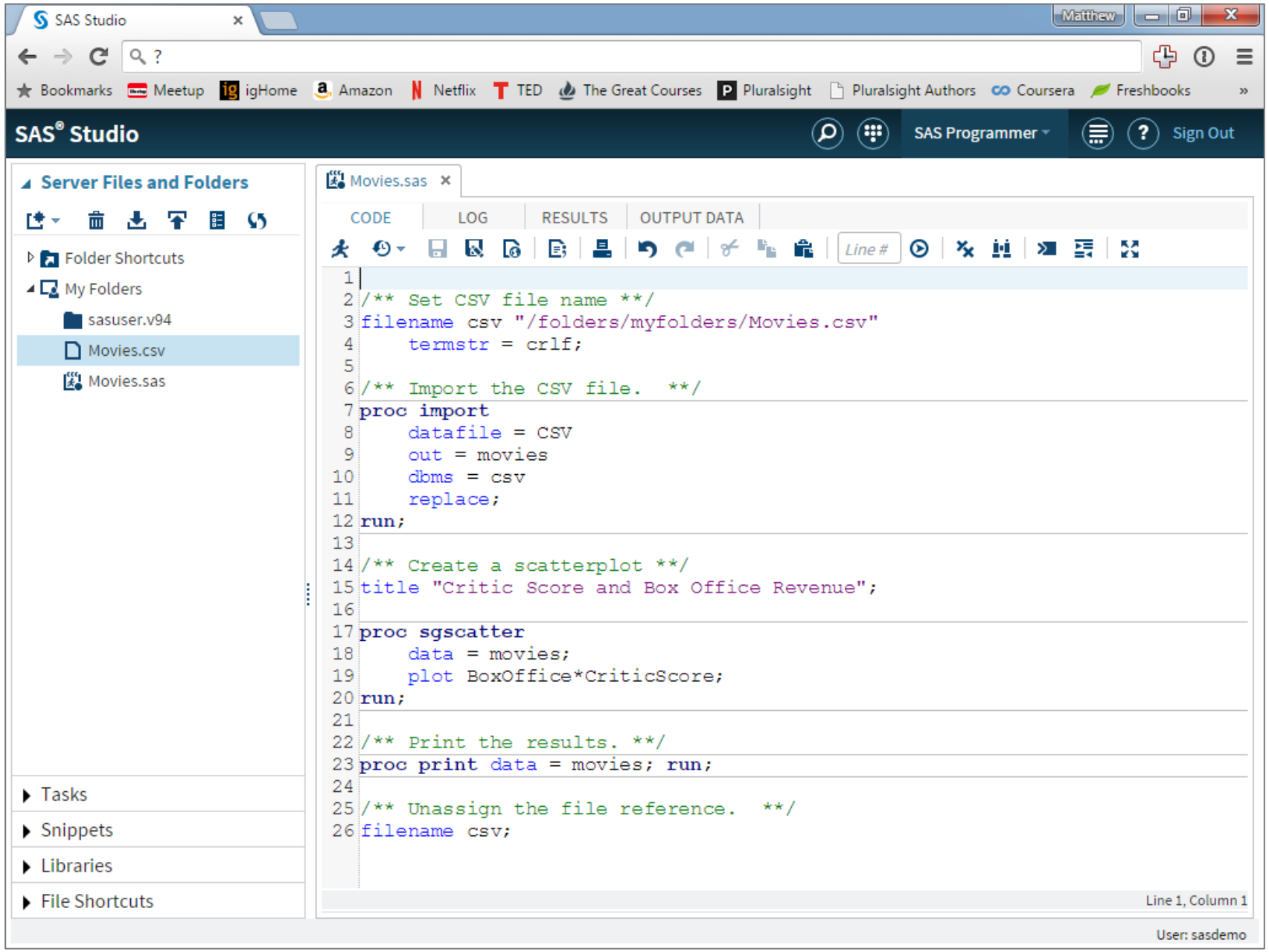

Movies.sav [DataSet2] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Reports
Descriptive Statistics
Custom Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Simulation...
Quality Control
ROC Curve...
Spatial and Temporal Modeling...
Automatic Linear Modeling...
Linear...
Curve Estimation...
Partial Least Squares...
Binary Logistic...
Multinomial Logistic...
Ordinal...
Probit...
Nonlinear...
Weight Estimation...
2-Stage Least Squares...
Optimal Scaling (CATREG)...
Visible: 6 of 6 Variables
Title
Runtime
Critic.Score
Box.Office
var
var
1 The Whole Nine Yards 98 45 57.3000
2 Gladiator 155 76 187.3000
3 Cirque du Soleil 39 45 13.4000
4 Dinosaur 82 65 135.6000
5 Big Momma's House 30 .5000
6 Gone in Sixty Seconds 24 101.0000
7 Shaft 68 70.3000
8 Me, Myself & Irene 48 .6000
9 Chicken Run 97 106.6000
10 The Patriot 62 2.3000
11 The Adventures of Rocky & B 43 .4000
12 The Kid 49 68.5000
13 X-Men 82 156.2000
14 What Lies Beneath 46 155.0000
15 Thomas and the Magic Railro 19 2.0000
16 Nutty Professor II: The Klump 26 122.4000
17 Hollow Man 27 72.1000
18 Coyote Ugly 22 59.8000
19 Space Cowboys 130 78 89.4000
20 The Replacements 118 40 44.0000
21 Bless the Child 107 3 28.1000
22 The Original Kings of Comedy 115 86 37.7000
23 The Cell 107 45 60.0000
24 The Crew 2000 PG-13 88 20 12.5000
25 Bring It On 2000 PG-13 98 63 67.9000
Data View
Variable View
Linear...
IBM SPSS Statistics Processor is ready
Cases: 100
Unicode:ON

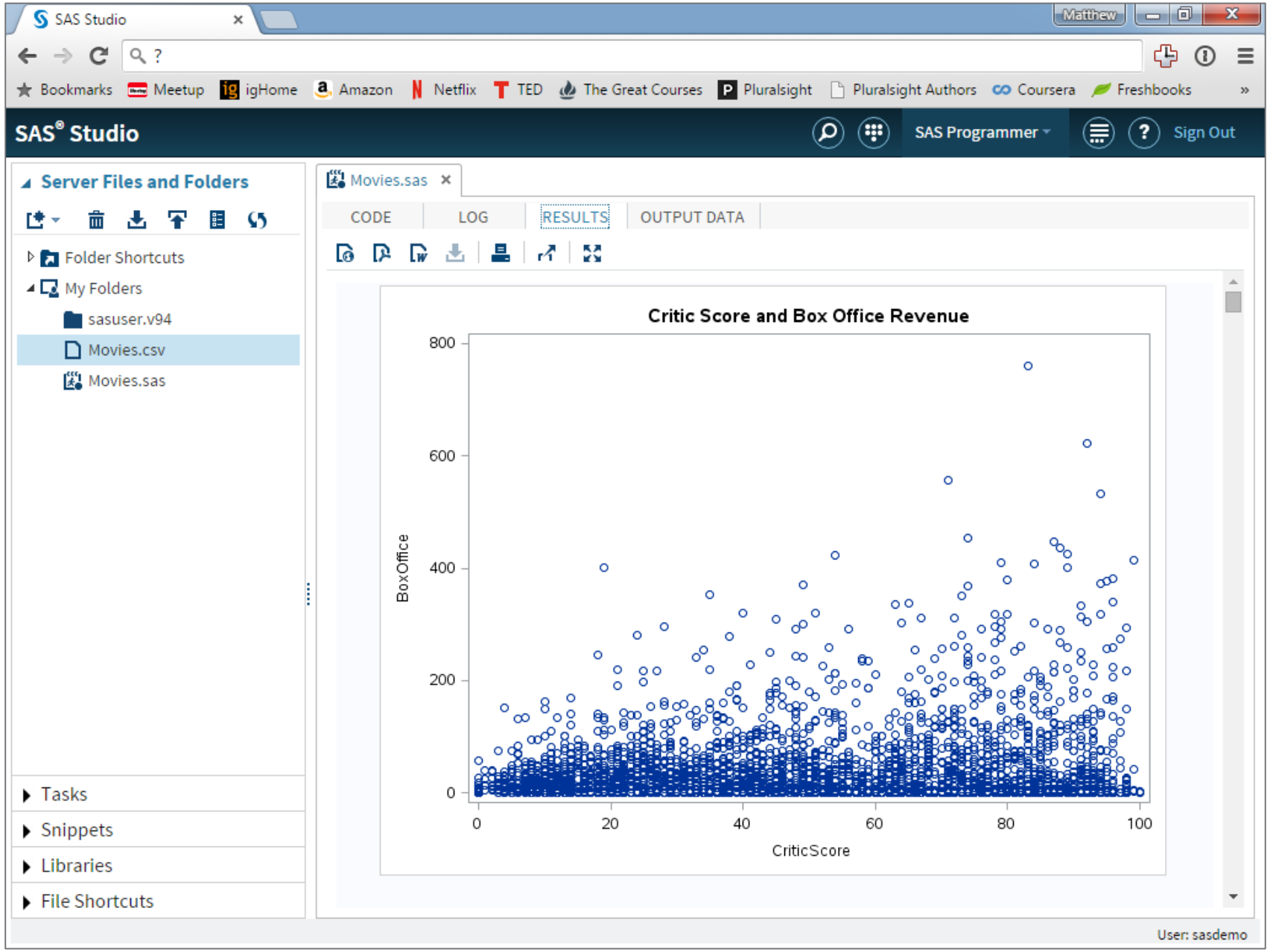

Movies.sav [DataSet2] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Visible: 8 of 8 Variables
Title Year Rating Runtime Critic.Score Box.Office Awards
1 The Whole Nine Yards
2 Cirque du Soleil: Journey of Ma
3 Gladiator
4 Dinosaur
5 Big Momma's House
6 Gone in Sixty Seconds
7 Shaft
8 Chicken Run
9 Me, Myself & Irene
10 The Patriot
11 The Adventures of Rocky & Bu
12 The Kid
13 X-Men
14 What Lies Beneath
15 Thomas and the Magic Railroa
16 Nutty Professor II: The Klumps
17 Coyote Ugly
18 Hollow Man
19 Space Cowboys
20 The Replacements
21 Bless the Child
22 The Cell
23 The Original Kings of Comedy
24 The Art of War
25 The Crew
Data View
Variable View
Chart Builder
Variables:
Title
Year
Rating
Runtime
Critic.Score
Box.Office
Awards
International
No categories (scale variable)
Chart preview uses example data
Box.Office
Runtime
Gallery
Basic Elements
Groups/Point ID
Titles/Footnotes
Element Properties...
Options...
Choose from:
Favorites
Bar
Line
Area
Pie/Polar
Scatter/Dot
Histogram
High-Low
Boxplot
Dual Axes
OK
Paste
Reset
Cancel
Help
Unicode:ON

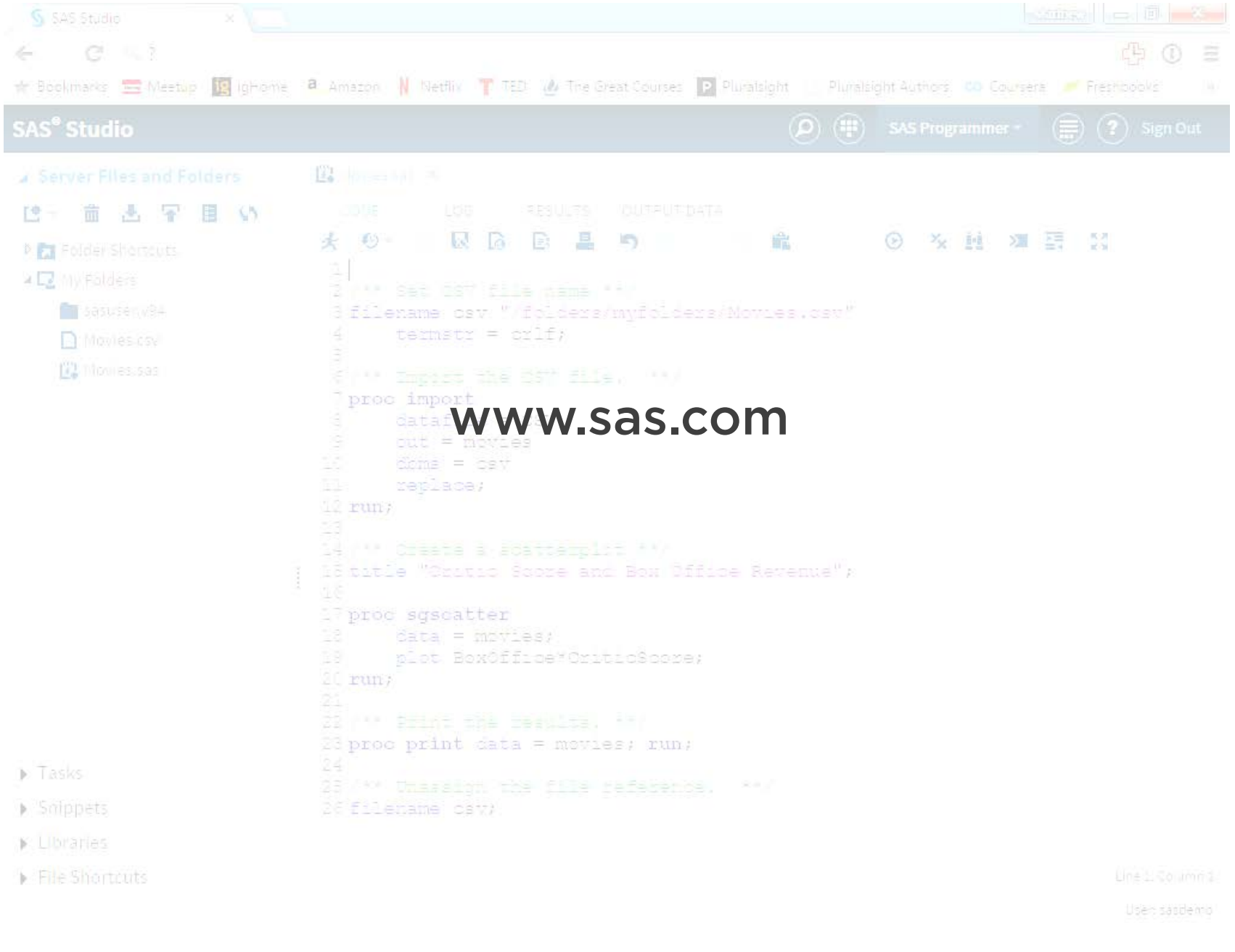

*Output2 [Document2] - IBM SPSS Statistics Viewer
File Edit View Data Transform Insert Format Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Output
Log
GGraph
Title
Notes
Graph
END GPL.
GGraph
800.0000
600.0000
400.0000
200.0000
.0000
Box.Office
0
50
100
150
200
250
Runtime
IBM SPSS Statistics Processor is ready
Cases: 100
Unicode:ON

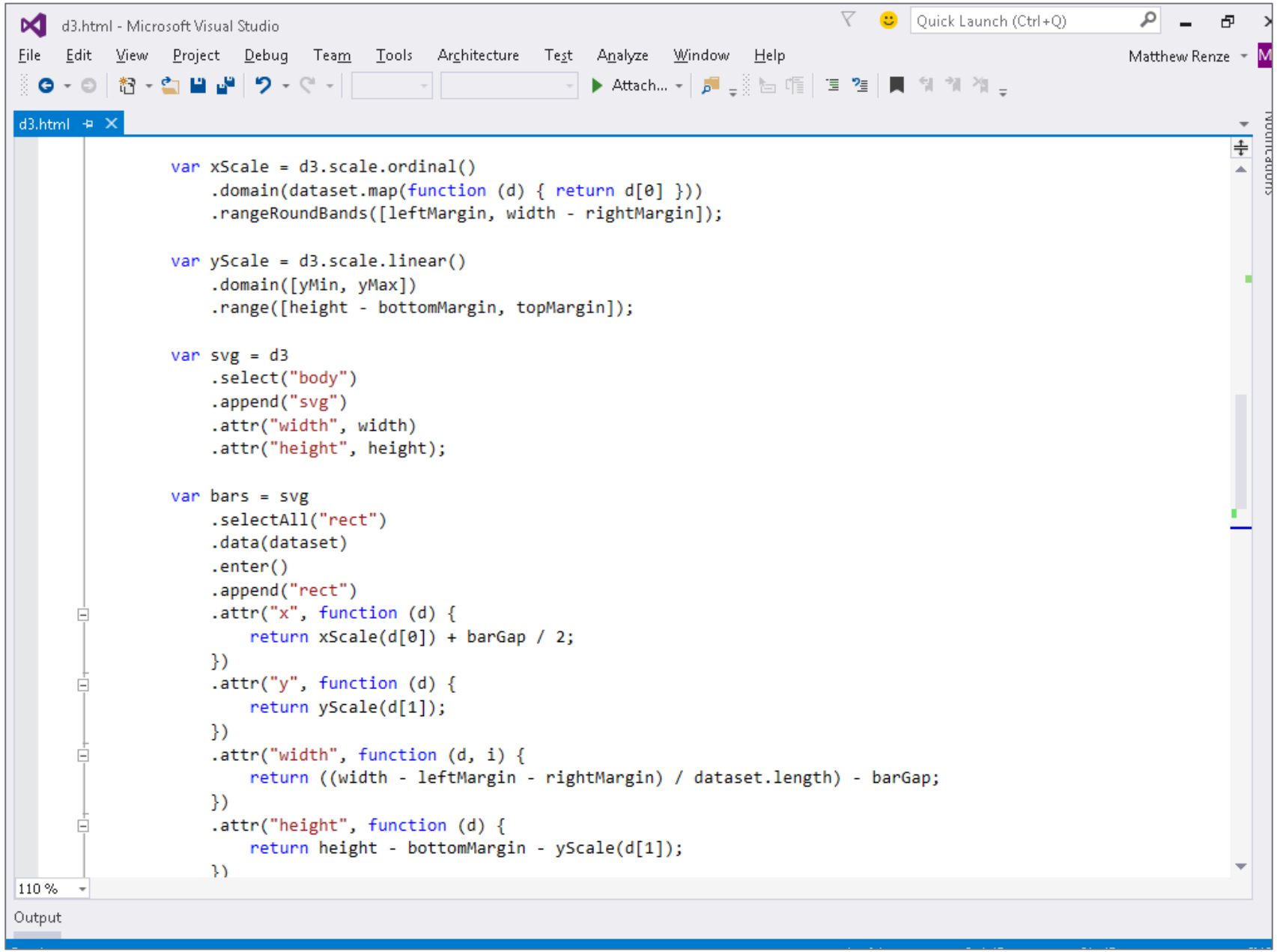


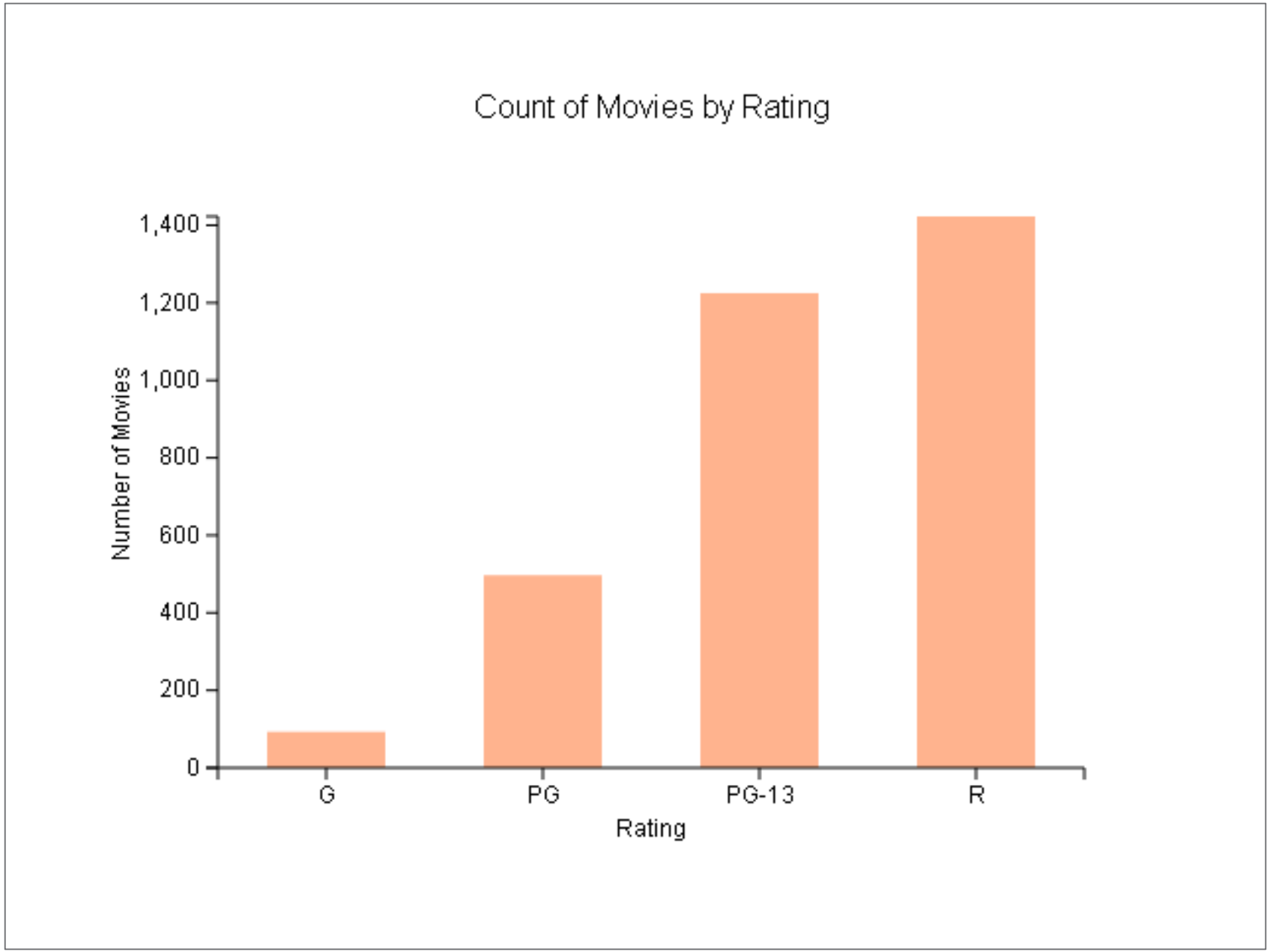
www.ibm.com/software/analytics/spss


SAS Studio
Matthew
Bookmarks
Meetup
igHome
Amazon
Netflix
TED
The Great Courses
Pluralsight
Pluralsight Authors
Coursera
Freshbooks
SAS® Studio
SAS Programmer
Sign Out
Server Files and Folders
Folder Shortcuts
My Folders
sasuser.v94
Movies.csv
Movies.sas
Tasks
Snippets
Libraries
File Shortcuts
Movies.sas
CODE
LOG
RESULTS
OUTPUT DATA
Line #
1
2 /** Set CSV file name **/
3 filename csv "/folders/myfolders/Movies.csv"
4 termstr = crlf;
5
6 /** Import the CSV file. **/
7 proc import
8 datafile = CSV
9 out = movies
10 dbms = csv
11 replace;
12 run;
13
14 /** Create a scatterplot **/
15 title "Critic Score and Box Office Revenue";
16
17 proc sgscatter
18 data = movies;
19 plot BoxOffice*CriticScore;
20 run;
21
22 /** Print the results. **/
23 proc print data = movies; run;
24
25 /** Unassign the file reference. **/
26 filename csv;
Line 1, Column 1
User: sasdemo


SAS Studio
Matthew
Bookmarks
Meetup
igHome
Amazon
Netflix
TED
The Great Courses
Pluralsight
Pluralsight Authors
Coursera
Freshbooks
SAS® Studio
SAS Programmer
Sign Out
Server Files and Folders
Folder Shortcuts
My Folders
sasuser.v94
Movies.csv
Movies.sas
Tasks
Snippets
Libraries
File Shortcuts
Movies.sas
CODE
LOG
RESULTS
OUTPUT DATA
Critic Score and Box Office Revenue
BoxOffice
CriticScore
0
200
400
600
800
0
20
40
60
80
100
User: sasdemo


SAS Studio
SAS® Studio
SAS Programmer
Sign Out
Server Files and Folders
Folder Shortcuts
My Folders
sasuser.v94
Movies.csv
Movies.sas
www.sas.com
Tasks
Snippets
Libraries
File Shortcuts


d3.html - Microsoft Visual Studio
Quick Launch (Ctrl+Q)
File Edit View Project Debug Team Tools Architecture Test Analyze Window Help
Matthew Renze
Attach...
d3.html
var xScale = d3.scale.ordinal()
.domain(dataset.map(function (d) { return d[0] }))
.rangeRoundBands([leftMargin, width - rightMargin]);
var yScale = d3.scale.linear()
.domain([yMin, yMax])
.range([height - bottomMargin, topMargin]);
var svg = d3
.select("body")
.append("svg")
.attr("width", width)
.attr("height", height);
var bars = svg
.selectAll("rect")
.data(dataset)
.enter()
.append("rect")
.attr("x", function (d) {
return xScale(d[0]) + barGap / 2;
})
.attr("y", function (d) {
return yScale(d[1]);
})
.attr("width", function (d, i) {
return ((width - leftMargin - rightMargin) / dataset.length) - barGap;
})
.attr("height", function (d) {
return height - bottomMargin - yScale(d[1]);
})
110 %
Output


Count of Movies by Rating
1,400
1,200
1,000
800
600
400
200
0
Number of Movies
G
PG
PG-13
R
Rating

d3.html - Microsoft Visual Studio

Quick Launch (Ctrl+Q)

File Edit View Project Debug Team Tools Architecture Test Analyze Window Help

Matthew Renze

Attach...

d3.html

```
var xScale = d3.scale.ordinal()
    .domain(dataset.map(function (d) { return d[0] }))
    .rangeRoundBands([leftMargin, width - rightMargin]);

var yScale = d3.scale.linear()
    .domain([yMin, yMax])
    .range([height - bottomMargin, topMar

var svg = d3
    .select("body")
    .append("svg")
    .attr("width", width)
    .attr("height", height);

var bars = svg
    .selectAll("rect")
    .data(dataset)
    .enter()
    .append("rect")
    .attr("x", function (d) {
        return xScale(d[0]) + barGap / 2;
    })
    .attr("y", function (d) {
        return yScale(d[1]);
    })
    .attr("width", function (d, i) {
        return ((width - leftMargin - rightMargin) / dataset.length) - barGap;
    })
    .attr("height", function (d) {
        return height - bottomMargin - yScale(d[1]);
    })
```

110 %

Output

Where to Go Next

R website: http://www.cran.r-project.org

R Studio: http://www.rstudio.com

Pluralsight: http://www.pluralsight.com

Coursera: https://www.coursera.org

Flowing Data: http://flowingdata.com

Matthew Renze: http://www.matthewrenze.com

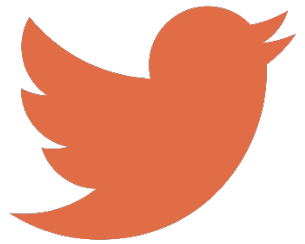
Feedback

Rate the course

Ask questions

Leave comments

Send a tweet



@matthewrenze

Course Summary

Introduction

Visualizing One Categorical Variable

Visualizing One Numeric Variable

Visualizing Two Categorical Variables

Visualizing Two Numeric Variables

Visualizing Both a Categorical and a Numeric Variable

Moving Beyond the Basics

Special Thanks

Basia Fusińska

Anne Herlache

Sarah Weno

Thank You!