CPSC203 WEEK-2 LAB-2
SPREADSHEET DESIGN RULES

-Prepared By
Nashad Ahmed Safa
Graduate Student
Department of Computer Science

SOME BASIC RULES ABOUT DESIGN

- Design the spreadsheet on paper first. Graph paper often works well.
- Test and edit your calculations. Where appropriate use intermediate calculations and check-sums to ensure calculations are correct.
- Keep the components of a calculation visible. No "magic numbers". Place fixed numbers used in a calculation in their own cell with a descriptive title.
- Be aware of the "space" or "geography" of the spreadsheet. Arrange your information so that it is well spaced and easy to take in at a glance.

PARTS OF A WELL DESIGNED SPREADSHEET

Adapted from: **The Elements of Spreadsheet Style. 1986. By J.M. Nevison**

These parts can be considered components of a disciplined approach to building spreadsheet so they are self-documenting. The parts could each be in their own sheet, or they could be in a single sheet.

- **Introduction** -- What is this spreadsheet or workbook about. Note the title, purpose, author, creation and revision dates etc.
- **Model and Assumptions** -- Justify any models, summary statistics, or calculated variables you are using.

PARTS OF A WELL DESIGNED SPREADSHEET

- **Data Dictionary** -- For every variable in the spreadsheet note: its
  - **Location** (cell range),
  - **Name**,
  - The **Data Class** it is (Raw Data, Statistical Summary, Calculated Variable, Score etc.),
  - **Data Type** (e.g. Integer, Text, Currency, Date, etc.) and
  - **Description** (a description of the data or what it's 'purpose' is).
- **Raw Data** -- Present your raw data in tabular form -- with **columns representing variables** and **rows representing cases**.

PARTS OF A WELL DESIGNED SPREADSHEET

- **Calculated Data**
  - Summary Statistics -- Usually Summary statistics result from calculations across rows for a single column.
  - Derived variables are often based on calculations across columns for a row.
- **Presentation** (Reporting)
  - Emphasize the final information you wish to show without excessive background details.
  - Use charts wherever appropriate to summarize large volumes of data.

EXAMPLE

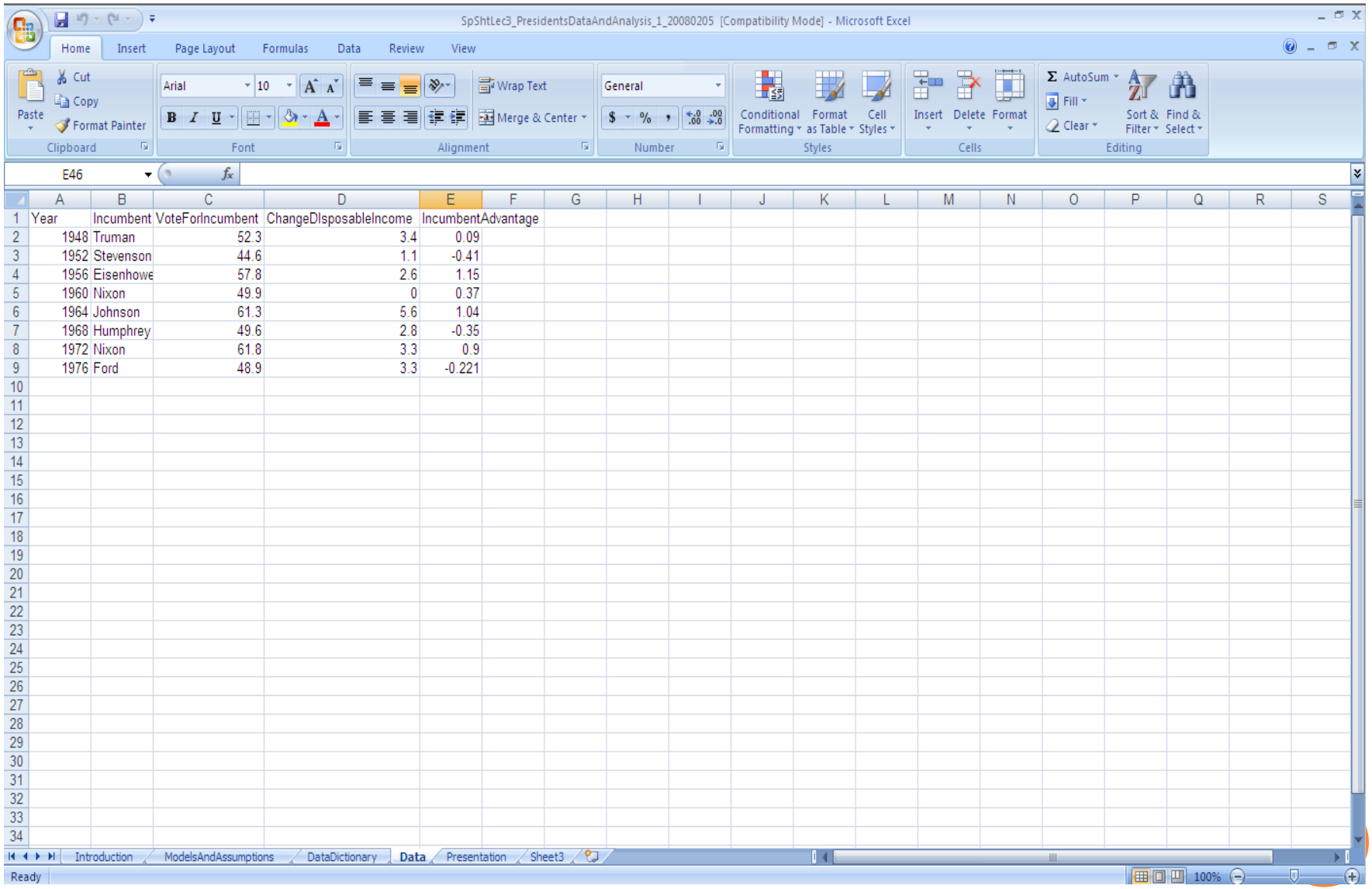
- The following example shows incremental improvements, working towards good spreadsheet design

Part 1

In this first iteration, two separate sheets are used:

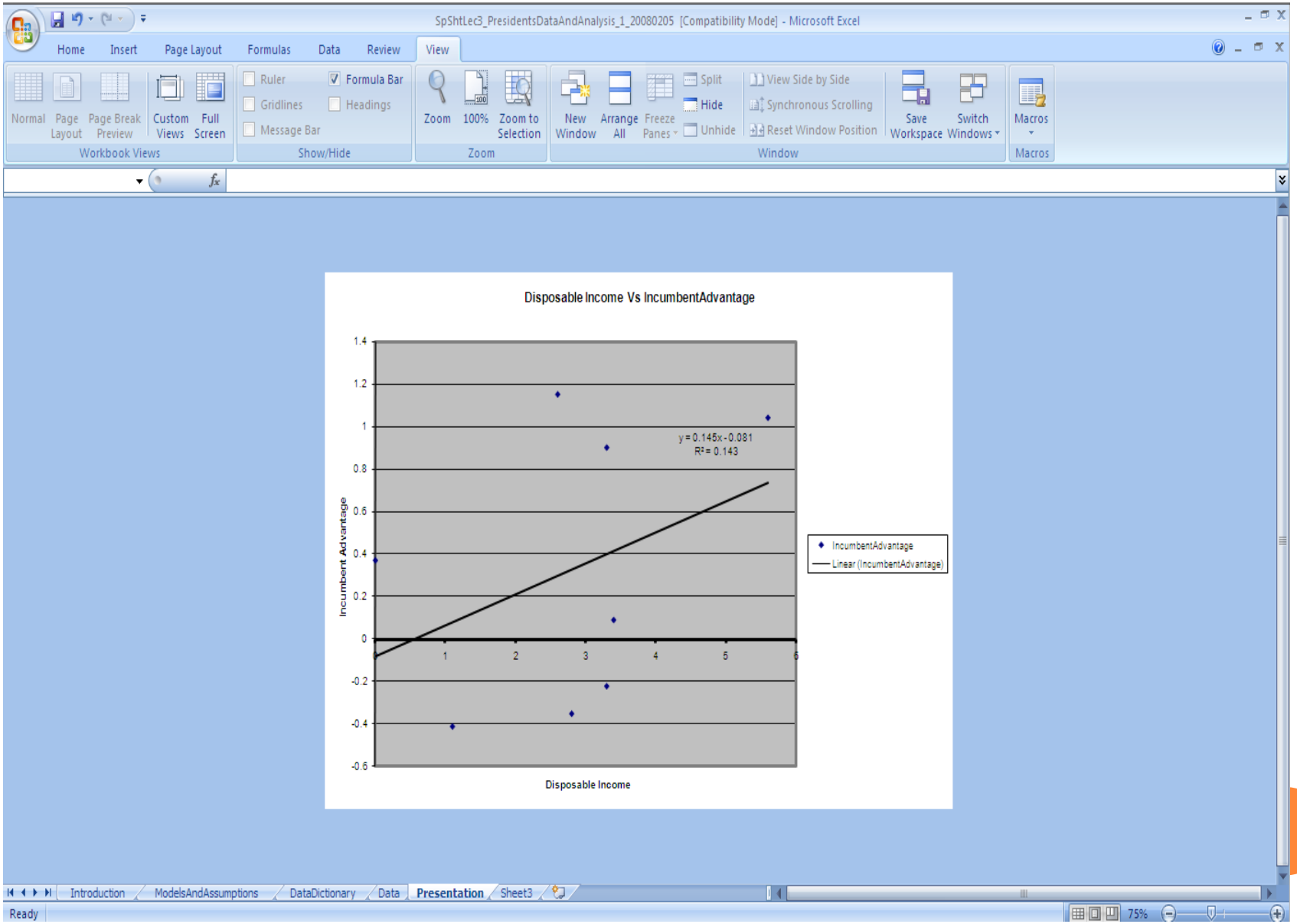
- data
- presentation

EXAMPLE



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Year	Incumbent	VoteForIncumbent	ChangeDIsposableIncome	IncumbentAdvantage														
2	1948	Truman	52.3	3.4	0.09														
3	1952	Stevenson	44.6	1.1	-0.41														
4	1956	Eisenhowe	57.8	2.6	1.15														
5	1960	Nixon	49.9	0	0.37														
6	1964	Johnson	61.3	5.6	1.04														
7	1968	Humphrey	49.6	2.8	-0.35														
8	1972	Nixon	61.8	3.3	0.9														
9	1976	Ford	48.9	3.3	-0.221														
10																			
11																			
12																			
13																			
14																			
15																			
16																			
17																			
18																			
19																			
20																			
21																			
22																			
23																			
24																			
25																			
26																			
27																			
28																			
29																			
30																			
31																			
32																			
33																			
34																			

EXAMPLE


SpShtLec3_PresidentsDataAndAnalysis_1_20080205 [Compatibility Mode] - Microsoft Excel
Home
Insert
Page Layout
Formulas
Data
Review
View
Normal
Page Layout
Page Break Preview
Custom Views
Full Screen
Workbook Views
Ruler
Gridlines
Message Bar
Formula Bar
Headings
Show/Hide
Zoom
100%
Zoom to Selection
New Window
Arrange All
Freeze Panes
Split
Hide
Unhide
View Side by Side
Synchronous Scrolling
Reset Window Position
Save Workspace
Switch Windows
Window
Macros
Disposable Income Vs IncumbentAdvantage
y = 0.145x - 0.081
R² = 0.143
Incumbent Advantage
Disposable Income
IncumbentAdvantage
Linear (IncumbentAdvantage)
Introduction
ModelsAndAssumptions
DataDictionary
Data
Presentation
Sheet3
Ready
75%

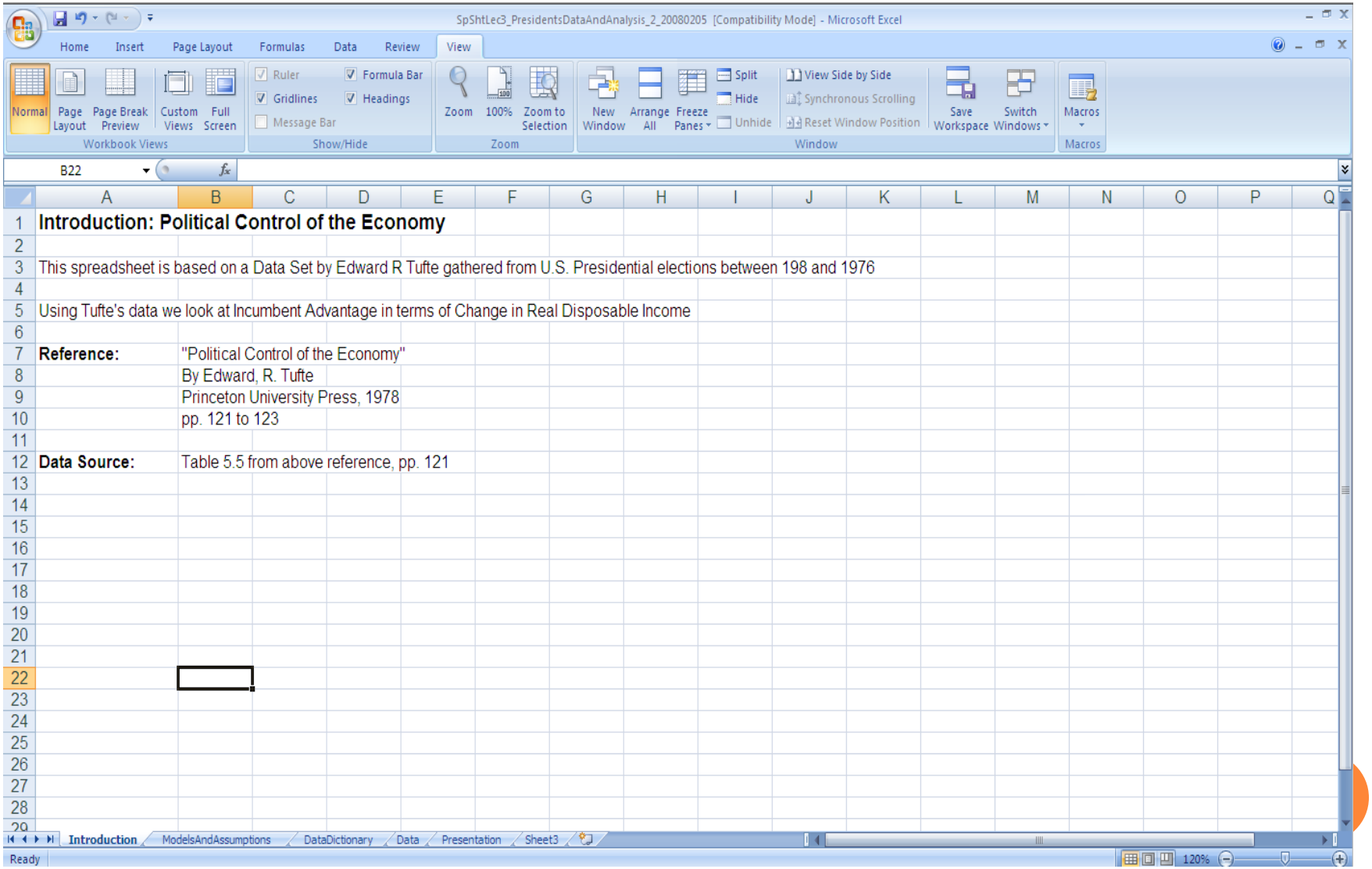
EXAMPLE

Part 2

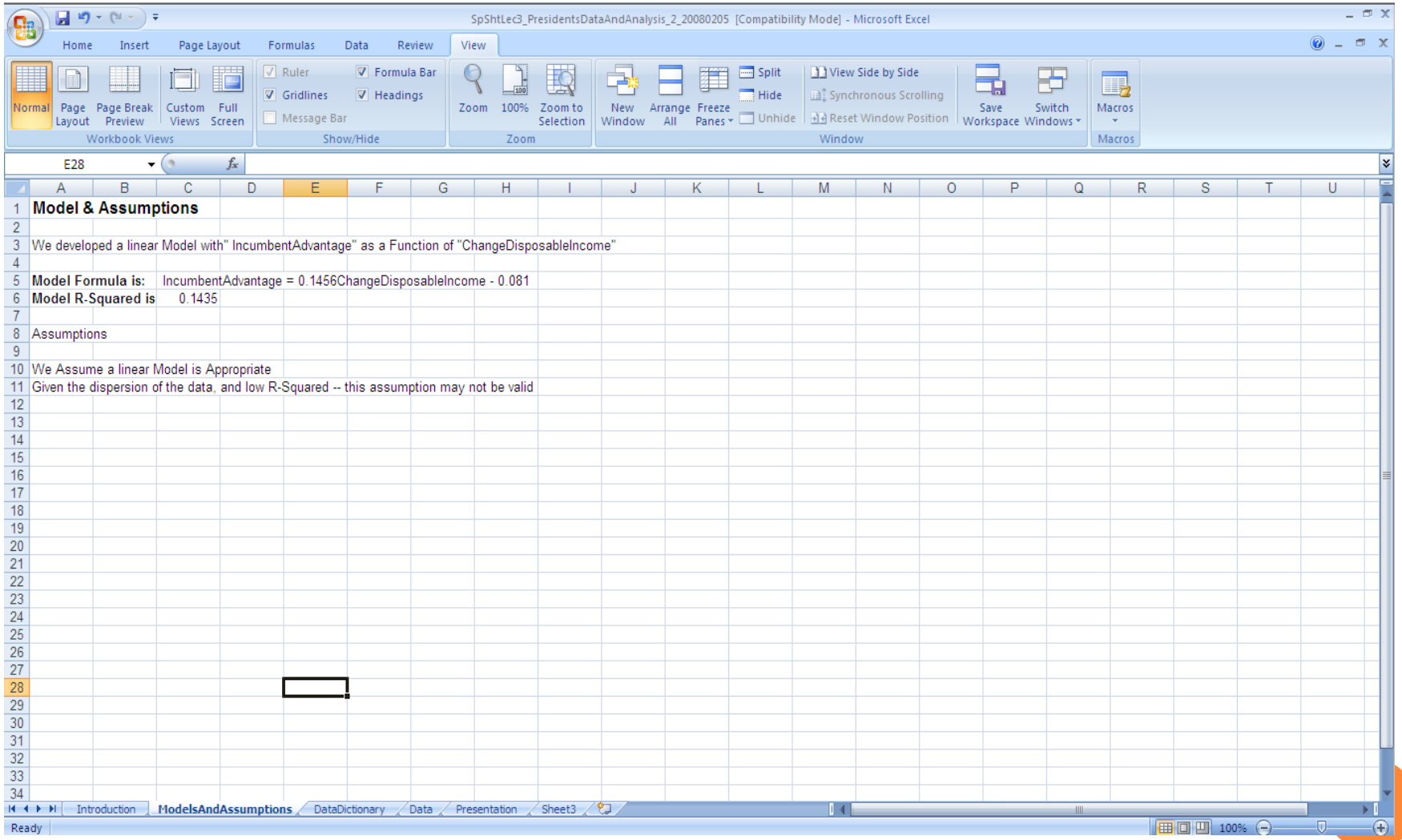
In this second iteration, five separate sheets are used:

- introduction
- model and assumptions
- data dictionary
- data
- presentation

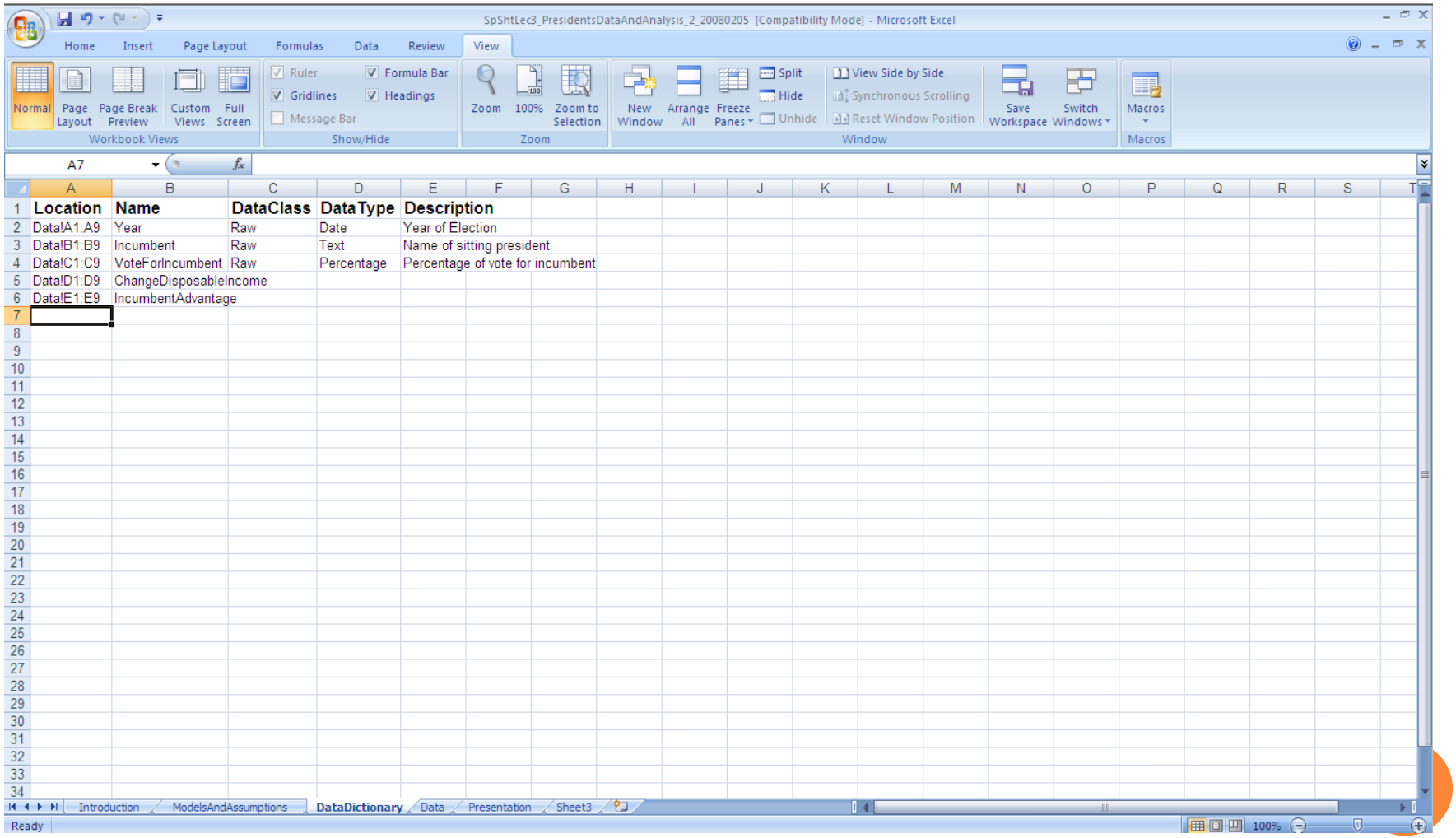
EXAMPLE


SpShtLec3_PresidentsDataAndAnalysis_2_20080205 [Compatibility Mode] - Microsoft Excel
Home
Insert
Page Layout
Formulas
Data
Review
View
Normal
Page Layout
Page Break Preview
Custom Views
Full Screen
Workbook Views
Ruler
Gridlines
Message Bar
Formula Bar
Headings
Show/Hide
Zoom
100%
Zoom to Selection
Zoom
New Window
Arrange All
Freeze Panes
Split
Hide
Unhide
View Side by Side
Synchronous Scrolling
Reset Window Position
Save Workspace
Switch Windows
Window
Macros
Macros
B22
Introduction: Political Control of the Economy
This spreadsheet is based on a Data Set by Edward R Tufte gathered from U.S. Presidential elections between 198 and 1976
Using Tufte's data we look at Incumbent Advantage in terms of Change in Real Disposable Income
Reference:
"Political Control of the Economy"
By Edward, R. Tufte
Princeton University Press, 1978
pp. 121 to 123
Data Source:
Table 5.5 from above reference, pp. 121
Introduction
ModelsAndAssumptions
DataDictionary
Data
Presentation
Sheet3
Ready
120%

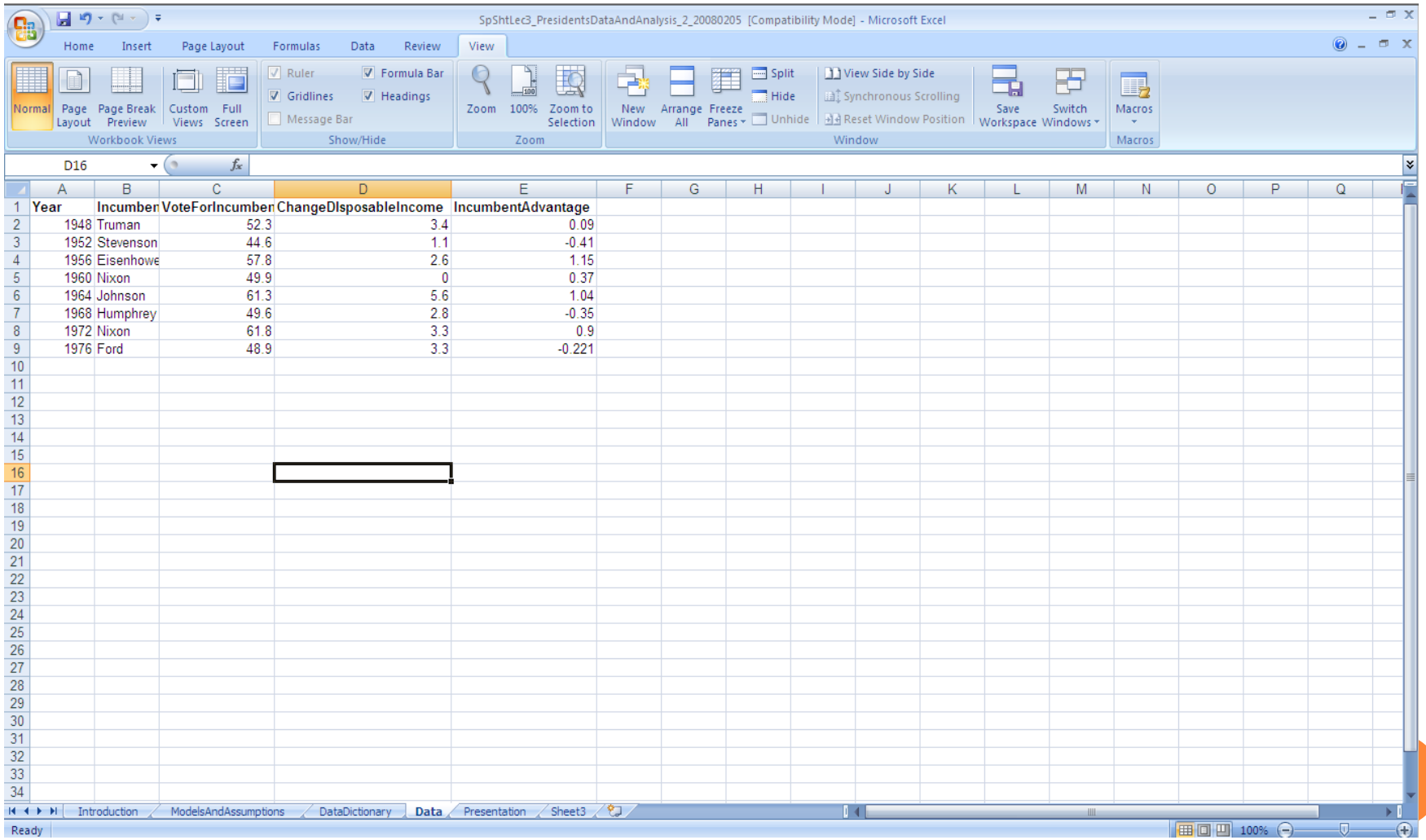
EXAMPLE


SpShtLec3_PresidentsDataAndAnalysis_2_20080205 [Compatibility Mode] - Microsoft Excel
Home
Insert
Page Layout
Formulas
Data
Review
View
Normal
Page Layout
Page Break Preview
Custom Views
Full Screen
Workbook Views
Ruler
Gridlines
Message Bar
Formula Bar
Headings
Show/Hide
Zoom
100%
Zoom to Selection
Zoom
New Window
Arrange All
Freeze Panes
Split
Hide
Unhide
View Side by Side
Synchronous Scrolling
Reset Window Position
Save Workspace
Switch Windows
Window
Macros
Macros
E28
Model & Assumptions
We developed a linear Model with" IncumbentAdvantage" as a Function of "ChangeDisposableIncome"
Model Formula is: IncumbentAdvantage = 0.1456ChangeDisposableIncome - 0.081
Model R-Squared is 0.1435
Assumptions
We Assume a linear Model is Appropriate
Given the dispersion of the data, and low R-Squared -- this assumption may not be valid
Introduction
ModelsAndAssumptions
DataDictionary
Data
Presentation
Sheet3
Ready
100%

EXAMPLE

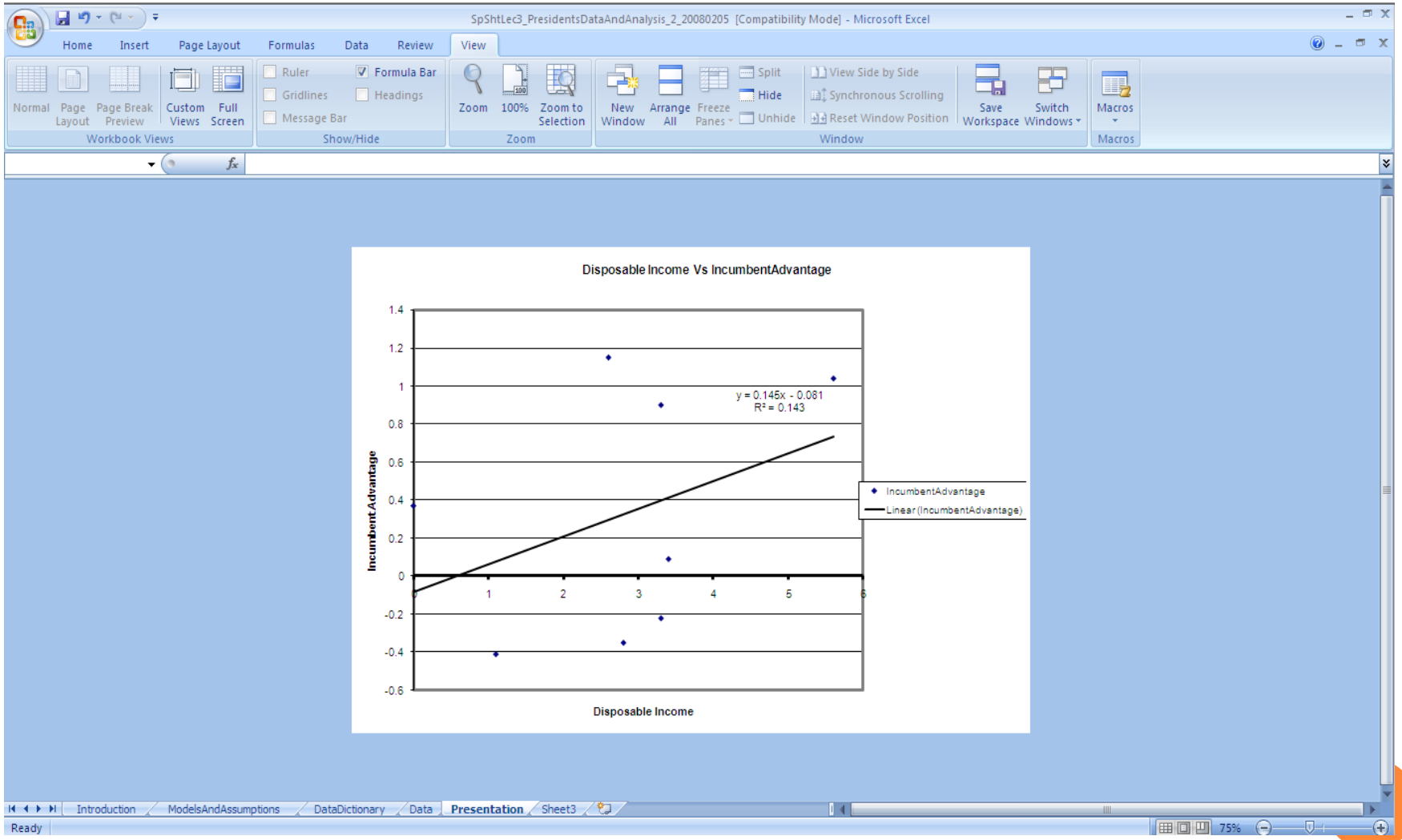

SpShtLec3_PresidentsDataAndAnalysis_2_20080205 [Compatibility Mode] - Microsoft Excel
Home
Insert
Page Layout
Formulas
Data
Review
View
Normal
Page Layout
Page Break Preview
Custom Views
Full Screen
Workbook Views
Ruler
Gridlines
Message Bar
Formula Bar
Headings
Show/Hide
Zoom
100%
Zoom to Selection
Zoom
New Window
Arrange All
Freeze Panes
Split
Hide
Unhide
View Side by Side
Synchronous Scrolling
Reset Window Position
Save Workspace
Switch Windows
Window
Macros
Macros
A7
Location
Name
DataClass
DataType
Description
Data!A1:A9
Year
Raw
Date
Year of Election
Data!B1:B9
Incumbent
Raw
Text
Name of sitting president
Data!C1:C9
VoteForIncumbent
Raw
Percentage
Percentage of vote for incumbent
Data!D1:D9
ChangeDisposableIncome
Data!E1:E9
IncumbentAdvantage
Introduction
ModelsAndAssumptions
DataDictionary
Data
Presentation
Sheet3
Ready
100%

EXAMPLE



	A	B	C	D	E
1	Year	Incumben	VoteForIncumber	ChangeDIsposableIncome	IncumbentAdvantage
2	1948	Truman	52.3	3.4	0.09
3	1952	Stevenson	44.6	1.1	-0.41
4	1956	Eisenhowe	57.8	2.6	1.15
5	1960	Nixon	49.9	0	0.37
6	1964	Johnson	61.3	5.6	1.04
7	1968	Humphrey	49.6	2.8	-0.35
8	1972	Nixon	61.8	3.3	0.9
9	1976	Ford	48.9	3.3	-0.221

EXAMPLE

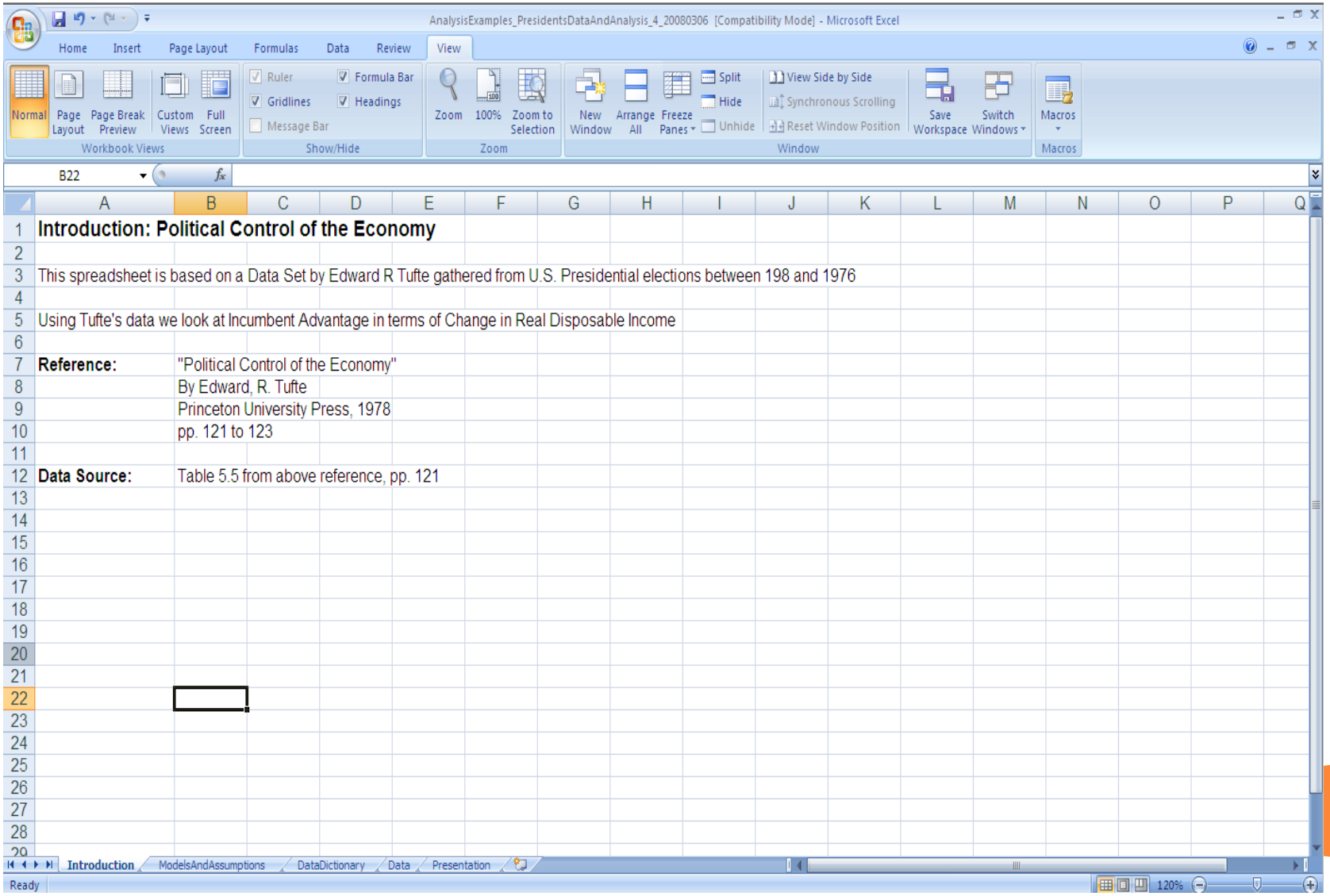

SpShtLec3_PresidentsDataAndAnalysis_2_20080205 [Compatibility Mode] - Microsoft Excel
Home
Insert
Page Layout
Formulas
Data
Review
View
Normal
Page Layout
Page Break Preview
Custom Views
Full Screen
Workbook Views
Ruler
Gridlines
Message Bar
Formula Bar
Headings
Show/Hide
Zoom
100%
Zoom to Selection
New Window
Arrange All
Freeze Panes
Split
Hide
Unhide
View Side by Side
Synchronous Scrolling
Reset Window Position
Save Workspace
Switch Windows
Window
Macros
Disposable Income Vs IncumbentAdvantage
y = 0.145x - 0.081
R² = 0.143
IncumbentAdvantage
Linear (IncumbentAdvantage)
Disposable Income
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
0
1
2
3
4
5
6
Introduction
ModelsAndAssumptions
DataDictionary
Data
Presentation
Sheet3
Ready
75%

EXAMPLE

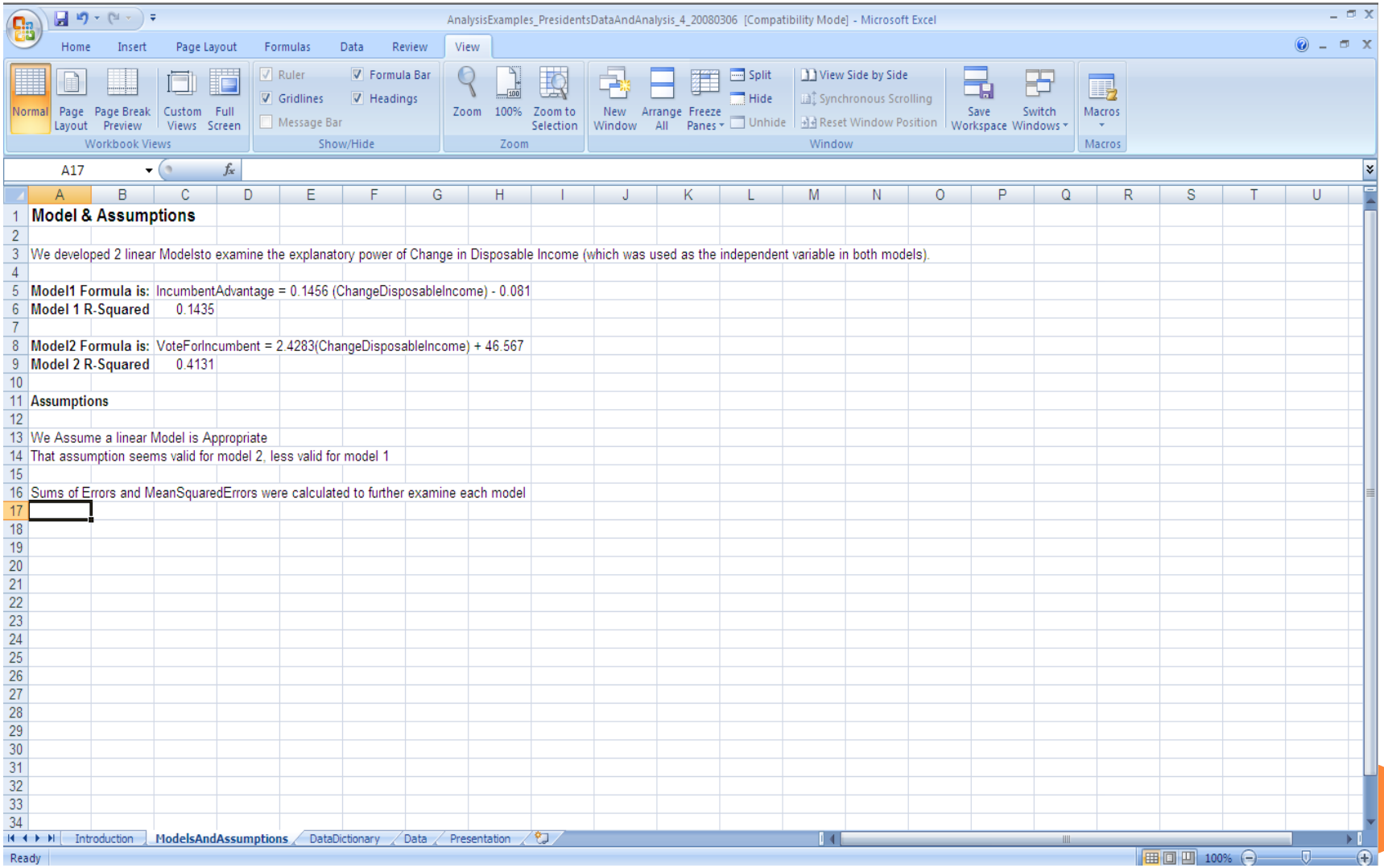
Part 3

- In this third iteration, the same five sheets are used (as in Part 2), with further refinements, descriptions, and presentation of the two models used.

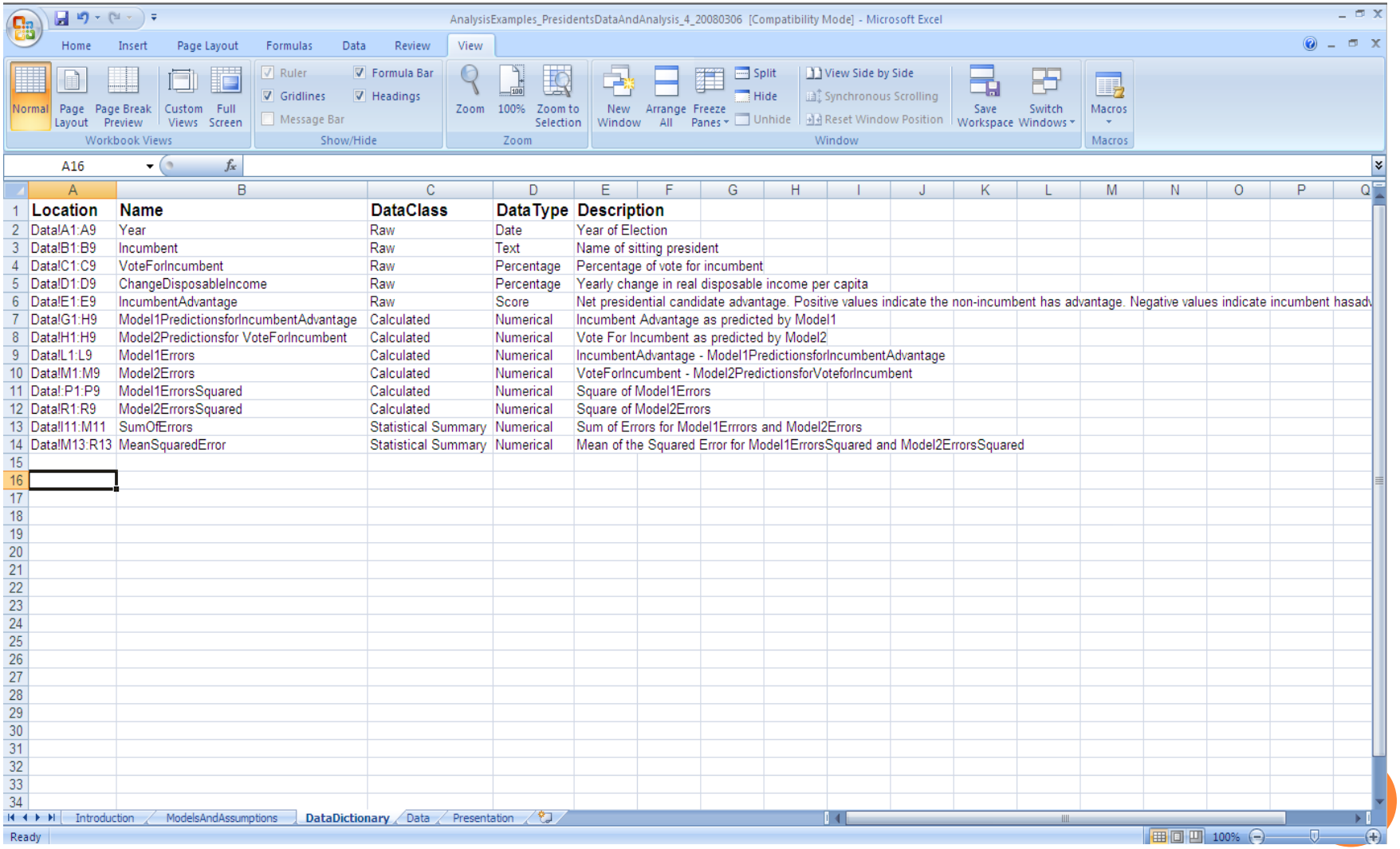
EXAMPLE


AnalysisExamples_PresidentsDataAndAnalysis_4_20080306 [Compatibility Mode] - Microsoft Excel
Home
Insert
Page Layout
Formulas
Data
Review
View
Normal
Page Layout
Page Break Preview
Custom Views
Full Screen
Workbook Views
Ruler
Gridlines
Message Bar
Formula Bar
Headings
Show/Hide
Zoom
100%
Zoom to Selection
Zoom
New Window
Arrange All
Freeze Panes
Split
Hide
Unhide
View Side by Side
Synchronous Scrolling
Reset Window Position
Save Workspace
Switch Windows
Window
Macros
Macros
B22
Introduction: Political Control of the Economy
This spreadsheet is based on a Data Set by Edward R Tufte gathered from U.S. Presidential elections between 198 and 1976
Using Tufte's data we look at Incumbent Advantage in terms of Change in Real Disposable Income
Reference:
"Political Control of the Economy"
By Edward, R. Tufte
Princeton University Press, 1978
pp. 121 to 123
Data Source:
Table 5.5 from above reference, pp. 121
Introduction
ModelsAndAssumptions
DataDictionary
Data
Presentation
Ready
120%

EXAMPLE

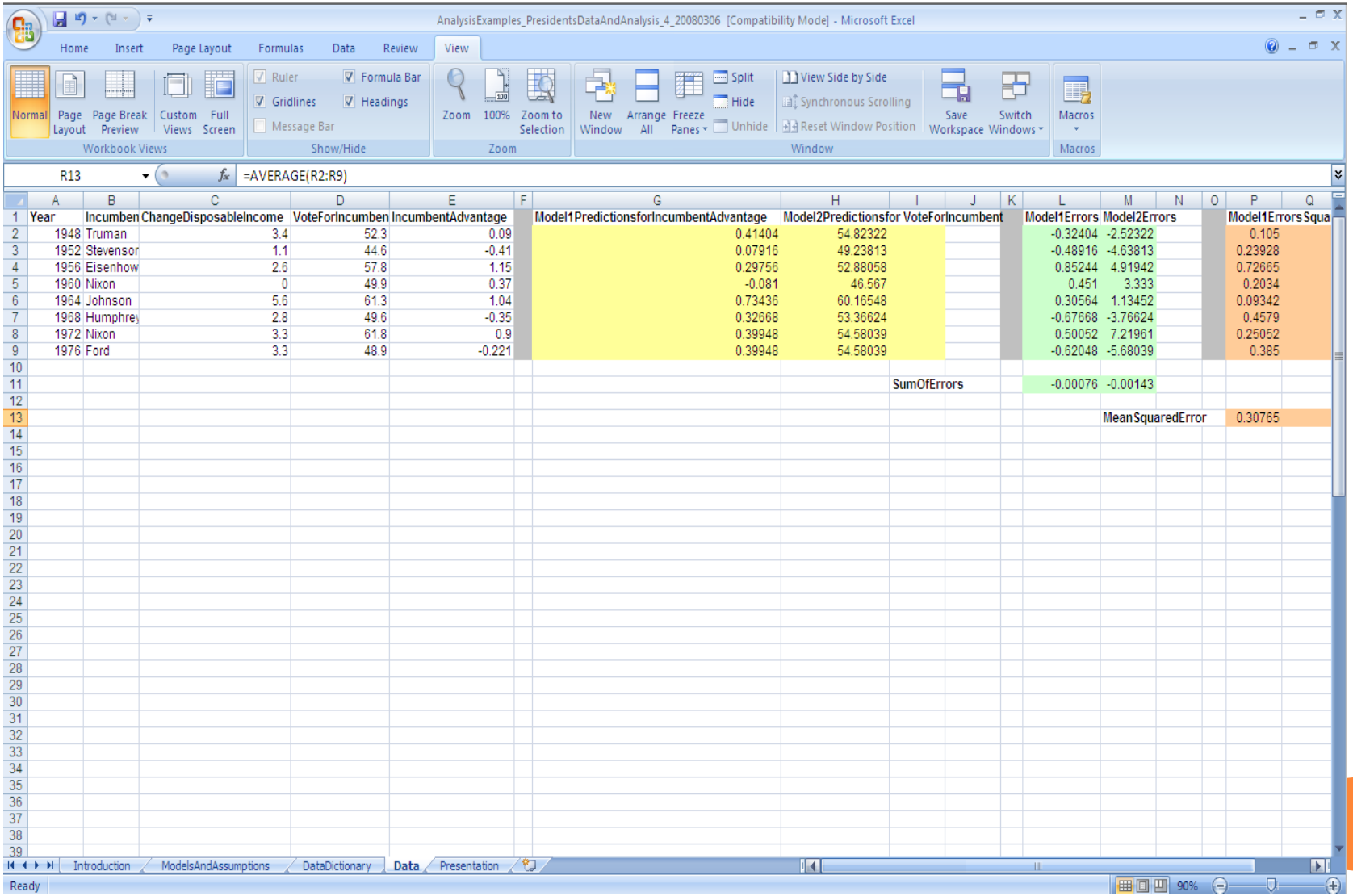

AnalysisExamples_PresidentsDataAndAnalysis_4_20080306 [Compatibility Mode] - Microsoft Excel
Home
Insert
Page Layout
Formulas
Data
Review
View
Normal
Page Layout
Page Break Preview
Custom Views
Full Screen
Workbook Views
Ruler
Gridlines
Message Bar
Formula Bar
Headings
Show/Hide
Zoom
100%
Zoom to Selection
New Window
Arrange All
Freeze Panes
Split
Hide
Unhide
View Side by Side
Synchronous Scrolling
Reset Window Position
Save Workspace
Switch Windows
Window
Macros
A17
Model & Assumptions
We developed 2 linear Modelsto examine the explanatory power of Change in Disposable Income (which was used as the independent variable in both models).
Model1 Formula is: IncumbentAdvantage = 0.1456 (ChangeDisposableIncome) - 0.081
Model 1 R-Squared 0.1435
Model2 Formula is: VoteForIncumbent = 2.4283(ChangeDisposableIncome) + 46.567
Model 2 R-Squared 0.4131
Assumptions
We Assume a linear Model is Appropriate
That assumption seems valid for model 2, less valid for model 1
Sums of Errors and MeanSquaredErrors were calculated to further examine each model
Introduction
ModelsAndAssumptions
DataDictionary
Data
Presentation
Ready
100%

EXAMPLE



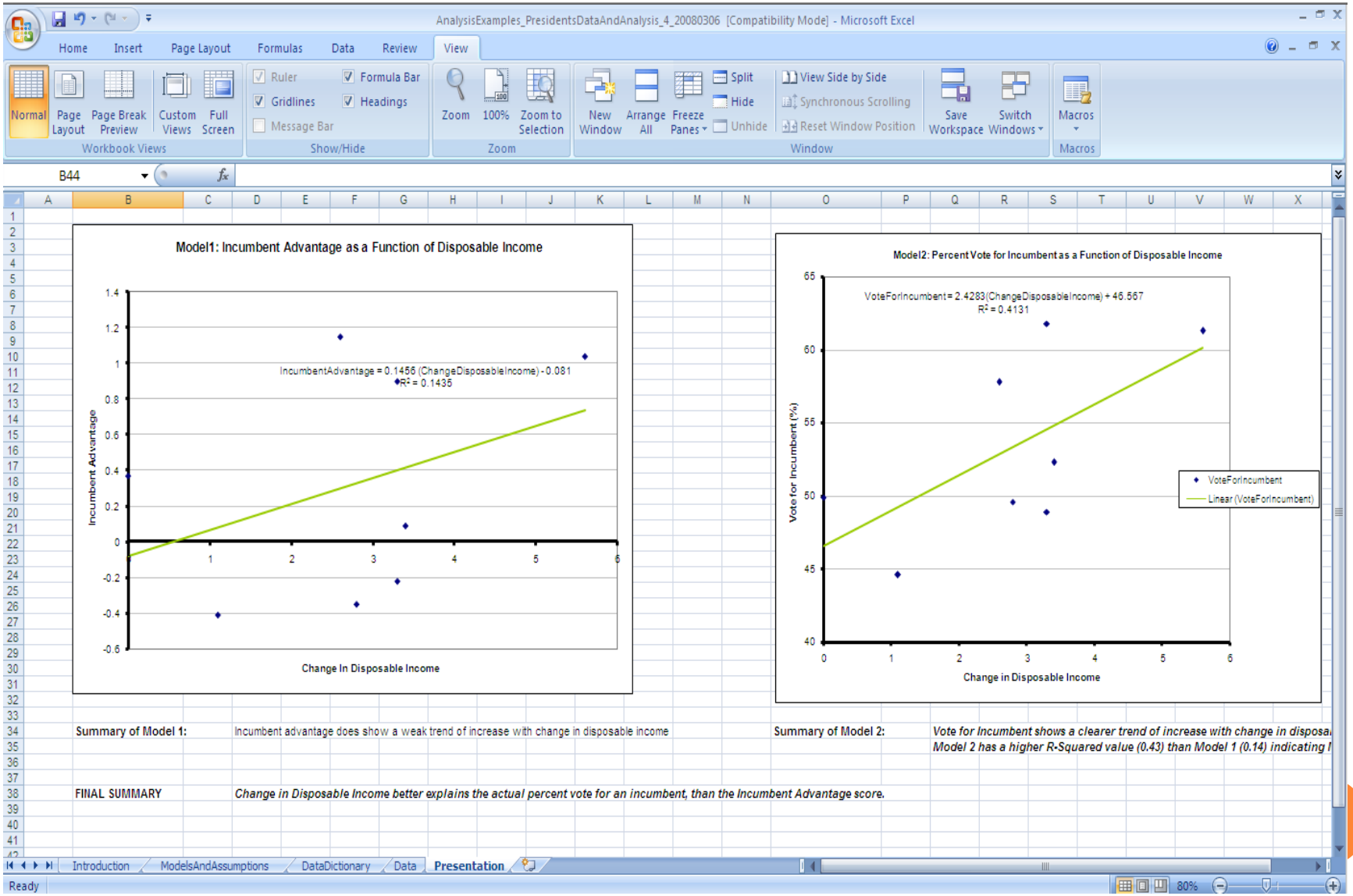
	A	B	C	D	E
1	**Location**	**Name**	**DataClass**	**DataType**	**Description**
2	Data!A1:A9	Year	Raw	Date	Year of Election
3	Data!B1:B9	Incumbent	Raw	Text	Name of sitting president
4	Data!C1:C9	VoteForIncumbent	Raw	Percentage	Percentage of vote for incumbent
5	Data!D1:D9	ChangeDisposableIncome	Raw	Percentage	Yearly change in real disposable income per capita
6	Data!E1:E9	IncumbentAdvantage	Raw	Score	Net presidential candidate advantage. Positive values indicate the non-incumbent has advantage. Negative values indicate incumbent hasadv
7	Data!G1:H9	Model1PredictionsforIncumbentAdvantage	Calculated	Numerical	Incumbent Advantage as predicted by Model1
8	Data!H1:H9	Model2Predictionsfor VoteForIncumbent	Calculated	Numerical	Vote For Incumbent as predicted by Model2
9	Data!L1:L9	Model1Errors	Calculated	Numerical	IncumbentAdvantage - Model1PredictionsforIncumbentAdvantage
10	Data!M1:M9	Model2Errors	Calculated	Numerical	VoteForIncumbent - Model2PredictionsforVoteforIncumbent
11	Data!:P1:P9	Model1ErrorsSquared	Calculated	Numerical	Square of Model1Errors
12	Data!R1:R9	Model2ErrorsSquared	Calculated	Numerical	Square of Model2Errors
13	Data!I11:M11	SumOfErrors	Statistical Summary	Numerical	Sum of Errors for Model1Errrors and Model2Errors
14	Data!M13:R13	MeanSquaredError	Statistical Summary	Numerical	Mean of the Squared Error for Model1ErrorsSquared and Model2ErrorsSquared

EXAMPLE



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Year	Incumben	ChangeDisposableIncome	VoteForIncumben	IncumbentAdvantage		Model1PredictionsforIncumbentAdvantage	Model2Predictionsfor	VoteForIncumbent			Model1Errors	Model2Errors			Model1ErrorsSqua	
2	1948	Truman	3.4	52.3	0.09		0.41404	54.82322				-0.32404	-2.52322			0.105	
3	1952	Stevensor	1.1	44.6	-0.41		0.07916	49.23813				-0.48916	-4.63813			0.23928	
4	1956	Eisenhow	2.6	57.8	1.15		0.29756	52.88058				0.85244	4.91942			0.72665	
5	1960	Nixon	0	49.9	0.37		-0.081	46.567				0.451	3.333			0.2034	
6	1964	Johnson	5.6	61.3	1.04		0.73436	60.16548				0.30564	1.13452			0.09342	
7	1968	Humphrey	2.8	49.6	-0.35		0.32668	53.36624				-0.67668	-3.76624			0.4579	
8	1972	Nixon	3.3	61.8	0.9		0.39948	54.58039				0.50052	7.21961			0.25052	
9	1976	Ford	3.3	48.9	-0.221		0.39948	54.58039				-0.62048	-5.68039			0.385	
10																	
11									SumOfErrors			-0.00076	-0.00143				
12																	
13													MeanSquaredError			0.30765	
14																	
15																	
16																	
17																	
18																	
19																	
20																	
21																	
22																	
23																	
24																	
25																	
26																	
27																	
28																	
29																	
30																	
31																	
32																	
33																	
34																	
35																	
36																	
37																	
38																	
39																	

EXAMPLE


AnalysisExamples_PresidentsDataAndAnalysis_4_20080306 [Compatibility Mode] - Microsoft Excel
Model1: Incumbent Advantage as a Function of Disposable Income
IncumbentAdvantage = 0.1456 (ChangeDisposableIncome) - 0.081
R² = 0.1435
Incumbent Advantage
Change In Disposable Income
Model2: Percent Vote for Incumbent as a Function of Disposable Income
VoteForIncumbent = 2.4283(ChangeDisposableIncome) + 46.567
R² = 0.4131
Vote for Incumbent (%)
Change in Disposable Income
VoteForIncumbent
Linear (VoteForIncumbent)
Summary of Model 1:
Incumbent advantage does show a weak trend of increase with change in disposable income
Summary of Model 2:
Vote for Incumbent shows a clearer trend of increase with change in disposa
Model 2 has a higher R-Squared value (0.43) than Model 1 (0.14) indicating
FINAL SUMMARY
Change in Disposable Income better explains the actual percent vote for an incumbent, than the Incumbent Advantage score.
Introduction
ModelsAndAssumptions
DataDictionary
Data
Presentation
Ready