Lecture Note #19: Data Analysis Tools Part #2

BUSI 201: Business Data Analysis

Topic 1. Manually Importing Data

Sometimes, you will be the one recording data in an Excel spreadsheet. But sometimes, you will be importing data from outside sources into Excel to perform data analysis. We can either manually import data ourselves, or rely on built-in tools that Excel has to offer. First, we will examine some basic manual data importing from outside sources.

Suppose you are interested secondary education attainment around the world. A quick search may lead you to a Wikipedia article titled "List of countries by secondary education attainment."¹ Figure 1 below is a screenshot of said webpage captured as of November 2023.

•••

The Free Encyclopedia	Q Search Wikipedia		Searc	h							
	List of countries b	y seco	ndary educa	ition	attai	nme	nt			文 _人 11	anguage 🗸
ents [hide]	Article Talk							Read	Edit V	'iew histor	ry Tools 🗸
1	From Wikipedia, the free encyclopedi	a						-			
f countries by percent ndary education attainment for ted age groups rences	This is a list of countries by the propo population of the relevant age groups sources.	rtion of the p that have o	population that has attain ompleted an upper second	ied at lea: ndary edu	st a second loation in th	dary educ ne listed o	ation. The countries. T	list is con 'he lists a	nposed of t re compile	he percer d from se	nt of the veral
	List of countries by perce	nt secor	idary education :	attainr	nent fo	r selec	ted age	group	OS [edit]		
	Country	♦ Year ♦	3 to 5 Years above graduation age ↓ (%) ^[1]	Year ¢	20–24 (%) ^[2] ¢	Year ¢	20–29 (%) ^[3] ≑	Year ¢	25–29 (%) ¢	Year ¢	25–34 (%) ^[4] ≑
	Country	 Year \$ 2014 	3 to 5 Years above graduation age ↓ (%) ^[1] 99	Year ¢	20–24 (%) ^[2] ¢	Year •	20–29 (%) ^[3] ¢ 98	Year ¢	25–29 (%) ¢	Year ¢	25–34 (%) ^[4] ≑
	Country South Korea Georgia	 Year 2014 2013 	3 to 5 Years above graduation age ¢ (%) ^[1] 99 96	Year ¢	20–24 (%) ^[2] ≑	Year ↓ 2015 2013	20-29 (%) ^[3] ♦ 98 95	Year ¢	25–29 (%) \$	Year \$	25–34 (%) ^[4] ≑
	Country (*) South Korea (*) Georgia • Japan	 Year 2014 2013 2016 	3 to 5 Years above graduation age ∳ (%) ^[1] 99 96 95	Year ¢	20–24 (%) ^[2] ¢	Year 2015 2013	20-29 (%) ^[3] ¢ 98 95	Year 🔶	25–29 (%) ¢	Year ¢	25–34 (%) ^[4] ≑
	Country South Korea Georgia Japan Croatia	 Year + 2014 2013 2016 2013 	3 to 5 Years above graduation age (%) ^[1] 99 96 95 95	Year	20–24 (%) ^[2] ◆ 95.7	Year 2015 2013 2013	20-29 (%) ^[3] ♦ 98 95 95	Year ¢	25–29 (%) ◆	Year 🔶	25–34 (%) ^[4] ¢
	Country South Korea Georgia Japan Crostia Ukraine	 Year \$ 2014 2013 2016 2013 2013 2012 	3 to 5 Years above graduation age (%) ^[1] 99 96 95 95 95	Year	20–24 (%) ^[2] ♦ 95.7	Year	20-29 (%) ^[3] ¢ 98 95 95 95 95 94	Year ¢	25–29 (%) ♦	Year ¢	25–34 (%) ^[4] ♦
	Country South Kores Georgia Japan Croatia Survaine Sweden	 Year \$ 2014 2013 2016 2013 2012 2013 	3 to 5 Years above graduation age (%) ^[1] ● 99 96 95 95 95 95 92 92	Year	20–24 (%) ^[2] ◆ 95.7 87.3	Year 2015 2013 2013 2013 2012 2012	20-29 (%) ^[3] ↓ 98 95 95 95 94 94	Year ¢	25–29 (%) ¢	Year	25-34 (%) ^[4] ♦
	Country South Korea Georgia Japan Croatia Ukraine Ukraine Fileweden I Ireland	 Year + 2014 2013 2016 2013 2012 2013 	3 to 5 Years above graduation age (%) ^[1] ● 99 96 95 95 95 95 92	Year ∳ 2015 2015 2015	20-24 (%) ^[2] ♦ 95.7 87.3 92.7	Year 2015 2013 2013 2013 2012 2012	20-29 (%) ^[3] ♦ 98 95 95 95 94 94	Year ¢	25–29 (%) ¢	Year ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓	25-34 (%) ^[4] ♦
	Country South Koree Georgia Japan Croatia Ukraine Sweden Kriand Armenia	 Year + 2014 2013 2016 2013 2012 2013 2013 2010 	3 to 5 Years above graduation age (%) ⁽¹⁾ 99 96 95 95 95 95 92 92	Year	20-24 (%) ^[2] ♦ 95.7 87.3 92.7	Year 2015 2013 2013 2012 2012 2012 2012 2010	20-29 (%) ⁽³⁾ ♦ 98 95 95 94 94 94 92	Year	25–29 (%) ♦	Year ↓ 2015 2015	25-34 (%) ^[4] ♦ 82 91
	Country South Korea Georgia Japan Croatia Ukraine Sweden Ikreland Kramenia Ukreland Kramenia Ukrel Kingdom	 Year + 2014 2013 2016 2013 2012 2013 2012 2013 2010 2010 2013 	3 to 5 Years above graduation age (%) ^[1] 999 966 955 955 925 92 93 93 94	Year ↓ 2015 2015 2015 2015 2015	20-24 (%) ^[2] ∳ 95.7 87.3 92.7 85.7	Year 2015 2013 2013 2012 2012 2012 2012 2010 2010 2013	20-29 (%) ³ ♦ 98 95 95 95 94 94 94 92 92	Year	25–29 (%) ♦	Year ↓ 2015 2015 2015 2015	25–34 (%) ^[4] ♦ 82 91 85
	Country South Korea Googia Japan Croatia Ukraine Sweden Ikraind Armenia Kutied Kingdom Kazakhstan	 Year + 2014 2013 2016 2013 2012 2013 2010 2010 2010 2010 2010 	3 to 5 Years above graduation age (%) ^[1] 99 96 95 95 95 95 95 95 92 1 93 94 93	Year ↓ 2015 2015 2015 2015 2015	20-24 (%) ^[2] ♦ 95.7 95.7 87.3 92.7 85.7	Year 2015 2013 2013 2012 2012 2012 2010 2010 2010	20-29 (%) ⁽³⁾ € 98 95 95 94 94 94 94 92 92 92	Year	25–29 (%) ♦	Year ↓ 2015 2015 2015 2015	25–34 (%) ^[4] ◆ 82 91 85
	Country South Kores Georgia Japan Croatia Croatia Kraine Kraine Kraine Kraine Kraine Kraine Country Country	 Year + 2014 2013 2016 2013 2012 2013 2010 2010 2010 2010 2010 2010 2010 2013 	3 to 5 Years above graduation age (%) ⁽¹⁾ 99 96 95 95 95 95 92 93 93 94 93 83	Year	20-24 (%) ^[2] ↓ 95.7 87.3 92.7 85.7 90.8	Year 2015 2013 2013 2012 2012 2012 2010 2010 2013 2010 2013	20-29 (%) ³ • 98 95 95 94 94 94 92 92 92 92 92	Year	25–29 (%) ↓	Year Year	25-34 (%) ^[4] ♦

Figure 1: Wikipedia Article

One way to import this data into Excel is to simply copy and paste the entire table. You can copy the data in the table by left clicking and dragging to select the table, and then right clicking the selected table, and selecting copy.

¹https://en.wikipedia.org/wiki/List_of_countries_by_secondary_education_attainment

You can choose two options when pasting the table data into Excel. You can choose to Keep Source Formatting, or Match Destination Formatting as shown in Figure 2. We will primarily be using the latter, as the source formatting is not necessarily well translated over to Excel.



Figure 2: Pasting Options

Pasting the table we copied earlier while matching destination formatting, we can import the table as shown in Figure 3. Now that we have the table in Excel, we can use the tools that we have at our disposal to "clean" the data. Remove redundant rows and columns, sorting data by educational attainment, color-coding the table using conditional formatting, creating charts to visualize data, etc.

	А	В	С	D	E	F	G	Н	1	J	K	L	М	N	0
1															
2															
3		Country	Year	3 to 5 Year	Year	20-24	Year	20-29	Year	25-29	Year	25-34			
4				graduation	n age	(%)[2]		(%) [3]		(%)		(%)[4]			
5				(%)[1]											
6		South Kor	2014	99			2015	98							
7		Georgia	2013	96			2013	95							
8		Japan	2016	95											
9		Croatia	2013	95	2015	95.7	2013	95							
10		Ukraine	2012	95			2012	94							
11		Sweden	2013	92	2015	87.3	2012	94			2015	82			
12		Ireland			2015	92.7					2015	91			
13		Armenia	2010	93			2010	92							
14		United Ki	2013	94	2015	85.7	2013	92			2015	85			
15		Kazakhsta	2010	93			2010	92							
16		Poland	2013	83	2015	90.8	2013	92			2015	94			
17		United St	2010	92			2013	91			2015	90			
18		Canada	2010	86			2010	91			2015	93			
19		Greece	2013	92	2015	89.6	2013	91			2015	84			
20		Slovakia	2013	93	2015	91.3	2013	90			2015	93			

Figure 3: Imported Table

This is a rather straightforward example of importing data. The original source material was already formatted as a table, and the importing process required little customization. Now, let us examine a case where the data requires a bit more work

CSV: Comma Separated Values

In some cases, you will encounter files in the form of pdfs or txt files. One such example can be found by downloading the BUSI201-LEC19-txt file. This file lists the top 20 movies of all time based on IMDB review scores as of November 2023.

```
BUSI201-LEC20-txt
                                 ×
                                       ^+
File
      Edit
             View
Title, Year, Length, IMDB Rating
The Shawshank Redemption, 1994, 142, 9.3
The Godfather, 1972, 175, 9.2
The Dark Knight, 2008, 202, 9
The Godfather Part II, 1974, 96, 9
12 Angry Men, 1957, 195, 9
Schindler's List, 1993, 201, 9
Pulp Fiction, 1994, 154, 8.9
The Lord of the Rings: The Fellowship of the Ring, 2001, 178, 8.8
The Good the Bad and the Ugly, 1966, 178, 8.8
Forrest Gump, 1994, 142, 8.8
Fight Club, 1999, 139, 8.8
The Lord of the Rings: The Two Towers, 2002, 179, 8.8
Inception, 2010, 148, 8.8
Star Wars: Episode V - The Empire Strikes Back, 1980, 124, 8.7
The Matrix, 1999, 136, 8.7
Goodfellas,1990,145,8.7
One Flew Over the Cuckoo's Nest,1975,143,8.7
Se7en,1995,127,8.6
It's a Wonderful Life,1946,130,8.6
```

Figure 4: Data in TXT Format

Copy the text file, and paste in into an Excel spreadsheet. The initial result will not be ideal, since each line in the text file will populate a single cell. We must call up the text import wizard by clicking **Paste Options**, and then **Use Text Import Wizard**.



Figure 5: Text File Pasted to Excel

The text import wizard pop-up is shown in Figure 6. Note that in the blue box that the source data is set to Delimited, since the entries are separated by commas. You should choose this same format when each field is separated by tabs as well.² This will be the default for most cases when you download a text file with data. Click Next to move along.

	Α	В	С	D	E	F	G	Н	1	J	K	L	М	N	0	Р	Q
1	Title,Year	Length, IN	IDB Rating														
2	The Shaws	hank Red	lemption,1	994,142,9.3	Text Imp	ort Wizar	d - Step 1 of	3					?	×			
3	The Godfat	ther,1972	,175,9.2		The Text V	/izard has	determined th	at vour data is	Delimited.								
4	The Dark K	night,200	8,202,9		If this is co	most show	and Next or ch	ooro the data t	me that hert	describes you	ur data						
5	The Godfat	ther Part	II,1974,96,9		II UIIS IS CO	rrect, criot	ise next, or ch	oose trie data tj	pe that best	uescribes you	il Gata.						
6	12 Angry M	len,1957,	195,9		Original	data type											
7	Schindler's	List,1993	3,201,9		Choose	the file typ	e that best des	scribes your da	ta:								
8	Pulp Fictio	n,1994,15	54,8.9		0	Delimited	 Character 	rs such as com	mas or tabs s	eparate each f	field.						
9	The Lord of	f the Ring	gs: The Fell	owship of the		Fixed <u>w</u> idt	h - Fields are	e aligned in colu	umns with sp	ices between	each field.						
10	The Good t	he Bad a	nd the Ugly	,1966,178,8.8													
11	Forrest Gu	mp,1994,	142,8.8		Start impo	rt at <u>r</u> ow:	1	÷ File orio	in: Windo	ws (ANSI)				~			
12	Fight Club,	1999,139	,8.8														
13	The Lord of	f the Ring	gs: The Two	Towers, 2002,	.1												
14	Inception,2	2010,148,	8.8		My da	ta has hea	ders.										
15	Star Wars:	Episode \	V - The Emp	oire Strikes Ba	c												
16	The Matrix	,1999,136	5,8.7		Preview	of selecte	d data:										
17	Goodfellas	,1990,14	5,8.7		1 1 1 1 1 1	e Veek	Tongth TM	D. Dating						- I			
18	One Flew (Over the	Cuckoo's N	est,1975,143,8	2 The	Shawsha	ink Redempt	tion, 1994, 1	42,9.3								
19	Se7en,1995	5,127,8.6			3 The	Godfath Dark Kr	her,1972,17	75,9.2									
20	It's a Wond	lerful Life	2,1946,130,	8.6	5 The	Godfath	er Part II	1,1974,96,9									
21					612 / 7 Schi	ndler's	n,1957,195 List,1993	5,9 3,201,9									
22														▶			
23													_				
24									Cancel	<	Back	Next >	Eini	ish			
25												_					
26																	

Figure 6: Text Import Wizard Step #1

In this next stage, we can tell Excel that the fields are separated by commas. So, in the red box of Figure 7, deselect Space, and select Comma. Observe how the preview in the blue box changes depending on the selected delimiters.

1	Α	В	С	D	E	E I	-	G	Н	1	J	К	L	М	N	0	Р	Q
1	Title,Year	Length,IN	1DB Rating															
2	The Shaws	hank Red	emption,19	994,142,9.3	3	Text Import	Wizard	- Step 2 of	3					?	×			
3	The Godfat	ther,1972	,175,9.2		1	This screen let	s vou set	the delimite	rs vour data o	ontains. You	can see how v	our text is a	ffected in the p	review below.				
4	The Dark K	night,200	8,202,9		1.0		/		,		,							
5	The Godfa	ther Part	1,1974,96,9			Delimiters												
6	12 Angry N	len,1957,	195,9			<u>T</u> ab												
7	Schindler's	5 List,1993	,201,9			Semicolo	on	✓ Treat	consecutive of	delimiters as o	ne							
8	Pulp Fictio	n,1994,15	4,8.9			Comma		Text qual	ifier "									
9	The Lord o	f the Ring	s: The Fello	owship of t	the I	Space		Text Stan										
10	The Good t	the Bad a	nd the Ugly	,1966,178,	8.8	Other:												
11	Forrest Gu	mp,1994,	142,8.8															
12	Fight Club,	1999,139,	8.8															
13	The Lord o	f the Ring	s: The Two	Towers,20	002,1													
14	Inception,	2010,148,	8.8															
15	Star Wars:	Episode \	/ - The Emp	ire Strikes	Bac	Data preview												
16	The Matrix	,1999,136	,8.7															
17	Goodfellas	5,1990,145	5,8.7			Title, Yea	ir,Lend	th, IMDB	Rating									
18	One Flew (Over the (Cuckoo's Ne	est,1975,14	43,8.	The			Shawshank	1020 125	Redem	otion,199	94,142,9.3					
19	Se7en,199	5,127,8.6				The			Dark	, 19/2, 1/3,	Knight	,2008,20	02,9					
20	It's a Wond	iertul Life	,1946,130,8	5.6		The 12			Godfather Angry		Part Men.19	57,195.9	9	II,1974,96	,9			
21					-	Schindle	's		List,1993	,201,9								
22						4												
23														_				
24										Cancel	<	Back	<u>N</u> ext >	Ein	ish			
20					-													

Figure 7: Text Import Wizard Step #2

 $^{^{2}}$ The tab key.

	А	В	С	D	E	F	G	Н	1	J	К	L	М	N	0	Р	Q
1	Title,Year	Length, IN	NDB Rating														
2	The Shaws	hank Rec	demption,19	994,142,9.3	3 Te	ext Import Wiza	rd - Step 3 of	3					?	×			
3	The Godfa	ther,1972	2,175,9.2		Thi	is screen lets vou	select each colu	umn and s	et the Data For	mat.							
4	The Dark K	(night,200	08,202,9			aluma data fami											
5	The Godfa	ther Part	II,1974,96,9			olumn data lonn	at										
6	12 Angry N	/len,1957,	,195,9			General		'Gene	eral' converts n	umeric values to	o numbers, dat	e values to da	ites, and all ren	naining			
7	Schindler's	s List,199	3,201,9		(Iext		value	s to text.					, in the second s			
8	Pulp Fictio	on,1994,1	54,8.9		(Date: MDY	\sim				Advanced						
9	The Lord o	of the Ring	gs: The Fello	owship of	thel	O Do not import	column (skip)										
10	The Good	the Bad a	nd the Ugly	,1966,178,	8.8												
11	Forrest Gu	imp,1994,	,142,8.8														
12	Fight Club	,1999,139	,8.8,														
13	The Lord o	of the Ring	gs: The Two	Towers,20	002,1												
14	Inception,	2010,148,	,8.8														
15	Star Wars:	Episode '	V - The Emp	ire Strikes	Bac D	ata <u>p</u> review											
16	The Matrix	x,1999,130	5,8.7											_ 11			
17	Goodfella	s,1990,14	5,8.7			General		Gener	GeneralGen	eral P. Pating							
18	One Flew	Over the	Cuckoo's Ne	est,1975,14	43,8.	The Shawshan	k Redemptio	n 1994	142 9.3	b Kacing							
19	Se7en,199	5,127,8.6				The Godfathe The Dark Kni	r aht	1972 2008	175 9.2 202 9								
20	It's a Wond	derful Life	e,1946,130,8	3.6		The Godfathe	r Part II	1974	96 9								
21						12 Angry Men Schindler's	List	1957 1993	195 9 201 9								
22					-	-								-			
23													_	_			
24									Cance	<	Back	Next >	Ein	ish			
25						_	_	_	_								
26																	

Figure 8: Text Import Wizard Step #3

In the third and last step of the text import wizard, users can set the data format for each column using the options in the green box. Then, check the preview in blue box, and select Finish. Figure 9 is the resulting table that is generated using the text import wizard.

	А	В	С	D	E	F	G	Н	l I	J	K	L
1	Title	Year	Length	IMDB Rati	ng							
2	The Shaws	1994	142	9.3								
3	The Godfa	1972	175	9.2								
4	The Dark I	2008	202	9								
5	The Godfa	1974	96	9								
6	12 Angry I	1957	195	9								
7	Schindler'	1993	201	9								
8	Pulp Fictic	1994	154	8.9								
9	The Lord o	2001	178	8.8								
10	The Good	1966	178	8.8								
11	Forrest Gu	1994	142	8.8								
12	Fight Club	1999	139	8.8								
13	The Lord o	2002	179	8.8								
14	Inception	2010	148	8.8								
15	Star Wars:	1980	124	8.7								
16	The Matri	1999	136	8.7								
17	Goodfella	1990	145	8.7								
18	One Flew	1975	143	8.7								
19	Se7en	1995	127	8.6								
20	lt's a Won	1946	130	8.6								
21					E							

Figure 9: Table Generated Using Text Import Wizard

Topic 2. Power Query: From Web

Instead of manually importing data, we can use Power Query to import, transform, and clean data. We will first examine how to directly import data from websites. Let us return to the Wikipedia article on secondary education attainment. Navigate to the Data tab, and select From Web. Type in the URL in the blue box, and then click OK.

File	Home	Inser	t Page L	ayout F	ormulas	Data Re	view Vie	w Autom	ate He	lp						
Get Data ~	Fro	om Text/C om Web om Table/I Get 8	SV 🔠 Range 🎦 Transform D	From Pictur Recent Sour Existing Cor ata	re ¥ rces nnections	Refresh All ~	Queries Propertie Workboo ueries & Con	& Connection es ok Links mections	5	Stocks	Currencies Data Types	Geography	4 ×	$ \begin{array}{c} $	Filter	Clear Reapply Advanced
A1	~) : [×	√ <i>f</i> x													
1	A	В	С	D	Е	F	G	Н	I.	J	К	L	М	N	0	Р
2 3 4 5 6 7 7 8 9 10 11 11 12 13 14		Fro	m Web sic ○ Advi	anced viki/List_of_	countries_b	y_secondary_	_education_a	attainment			ок с	ancel				

Figure 10: Importing from the Web

Once Excel establishes a link to the webpage, it will open up the Navigator page as shown in Figure 11. To the left hand side, you will find the objects included in the webpage. For this purpose, we should select TableO, and check out a preview of the table in the blue box. If the preview in the blue box indeed matches the table you wish to import, select Transform Data.

A	В	С	D	E	F	G	н	1.1	J.	К	L	M	N	0	Р	Q	R	S	Т	U	V	W	х	Y	Z	AA	AB	AC
1																												
2																			<)									
3					Nav	vigator																						
9					1101	igutor																						
6									\$		Table View	Web View	,															
7										- 1																		
8					L Se	ect multipl	le items				Table 0							là l										
9					Displa	ay Options	*		G	à	Country		Year	3 to 5 Years	above gradu	ation age (%)[1	Year2											
10					A 🖬	https://e	n.wikiped	lia.org/wiki/Lis	t_of_countries_	-	Country		Year	3 to 5 Years a	above gradua	tion age (%)	Year											
11						Dorum	ent				South Korea		2014	99				^										
12											Georgia		2013	96				11										
10						III Table U	, 				Japan		2016	95														
15											Croatia		2013	95			2015											
16											Ukraine		2012	95				11										
17											Sweden		2013	92			2015	11										
18											Ireland		nul	1			nw/ 2015											
19											Armenia		2010	93				11										
20											United Kined	om	2013	94			2015											
21											Kazakhstan		2010	93														
22											Poland		2013	83			2015											
23											United State		2010	92														
24											Canada		2010	85														
25											Greece		2013	92			2015											
27											Slovakia		2013	93			2015											
28											Cyprus		2013	93			2015											
29											Slovenia		2013	89			2015											
30											Israel		2012	88														
31											Russian Fede	ration	2013	87														
32											Czech Repub	lic	2013	90			2015	\sim										
33											1						>											
34										- LL	× .							_										
30																												
37															Load -	Transform (leta Can	cel										
38																	_											
														_														

Figure 11: Selecting the Data

A new window named Power Query Editor will pop up, which allows the users to edit the data before we import it to Excel. The most basic operations here will be operating on rows and columns.

File	• - Tabl Home	e 0 - Ira	Power Query Edi Instorm Add (itor Lolumn View												— (
Close & Load • Close	Refresh Preview •	Pr Ac	operties Ivanced Editor anage • C	Choose Remove olumns • Columns • Manage Columns	Keep Rem Rows • Row Reduce Row	× 2↓ A↓ S × S Sort	Split Column •	iroup By	Data Type: Text • Use First Row a J ₂ Replace Values Transform	s Hi	eaders 🝷	Merge Queries • Append Queries • Combine Files Combine	Mana Parame Parame	ige ters ▼	Data source settings Data Sources	New Source •	
Oueries [1]		(-	Data@ //"Countr		th ("Veno" to	no to	("2 to 5 Years also		maduation	and (%)[1]: type text)	TVene2"		*1 14		
Table 0			- 181 181 - Country	- dl. Yes	- di amen	y , cype cer		pe ce	all News		El addactor i	age (w)[1], type text);	(Teor 2) (ype ce.	·····	Query Settings	×
		1	Country	Year	3 to 5 Years graduation (above ge	uation age (%)[1]		Year	20	C 20-24 0-24 6)	Year		vc 20-29 (%)	×	✓ PROPERTIES Name Table 0	
		2	South Korea	2014	99				nu	a		null 2015		98	_	All Properties	
		3	Georgia	2013	96				nu	dV .		null 2013		95	_	A APPLIED STEPS	
		4	Japan	2016	95				nu	di		null	null			- Arree sters	
		5	Croatia	2013	95				2015	95	5.7	2013		95	_	Source	*
		6	Ukraine	2012	95				nu	dl 🛛		null 2012		94	_	X Changed Tupp	н
		7	Sweden	2013	92				2015	87	7.3	2012		94		v changed type	
		8	Ireland		null			null	2015	92	2.7		null				
		9	Armenia	2010	93				nu	dV .		null 2010		92			
		10	United Kingdom	2013	94				2015	85	5.7	2013		92	_		
		11	Kazakhstan	2010	93				nu	d)		null 2010		92			
		12	Poland	2013	83				2015	90	0.8	2013		92			
		13	United States	2010	92				nu	d)		null 2013		91			
		14	Canada	2010	86				nu	dl .		null 2010		91			
		15	Greece	2013	92				2015	85	9.6	2013		91			
		16	Slovakia	2013	93				2015	91	1.3	2013		90			
		17	Cyprus	2013	93				2015	94	4.3	2013		90			
		18	Slovenia	2013	89				2015	90	0.9	2013		90			
		19	Israel	2012	88				nu	al I		null 2012		89			
		20	Russian Federation	2013	87				nu	dV .		null 2013		88			
		21	Czech Republic	2013	90				2015	90	0.4	2013		88			
		22	Lithuania	2013	91				2015	90	0.9	2013		88			
		23	France	2013	83				2015	87	7.2	2013		88	~		
		24	<												>		

Figure 12: Power Query Editor

Editing Columns

First, to select the columns that are relevant, we can click **Choose Columns** button shown in the red box in Figure 13. Then, you can choose the columns that you would like to have included in the table that will be imported into Excel. You can "uncheck" the items in the blue box that you would not like to have imported.

×	• = Tabl	e 0 - F	Power Query E	ditor										×
File	Home	Irar	nstorm Add	I Column V	/iew									~ 🕜
Close & Load •	Refresh Preview *	Prc Adv Ma	operties vanced Editor nage -	Choose Rer Columns • Colu	move Keep Rem mns* Rows* Row	A A A A A A A A A A A A A A A A A A A	Choose Columns Choose the columns to keep	ype: Text ▼	×	Append Queries *	Manage Parameters *	Data source settings	New Source •	
ciose		quer	,	manage cola	inits includee nov	5 Sort		91		combine	rarameters	Duta Dources	new query	
Queries [1]		i 🖂	√ fx = T	able.TransformCol	lumnTypes(Data0,{{"Countr	/", type text	Rectange -		ic	on age (%)[1]", type text}, {	"Year2", type tex	d), 🔤 🗸	Query Settings	×
Table 0			A ^B C Country	▼ A ^{II} C Year	✓ ^{Al} C 3 to 5 Y	ars above gradu	Country		(%)	[2] ▼ ^{A®} C Year3	✓ A ⁰ C 20-2	9 (%) [3]		
		1	Country	Year	3 to 5 Years graduation a (%)	ibove ge	 ✓ Year ✓ 3 to 5 Years above grade ✓ Year2 	ation age (%)[1]		Year	20–29 (%)	^	A PROPERTIES Name Table 0	
		2	South Korea	2014	99		20-24 (%)[2]			null 2015	98		All Properties	
		3	Georgia	2013	96		✓ Year3			null 2013	95		A APPLIED STEPS	
		4	Japan	2016	95		20-29 (%) [3]			null	null		Saura	
		5	Croatia	2013	95		✓ Year4			2013	95	_	Navigation	8
		6	Ukraine	2012	95		25-29 (%)			null 2012	94		X Changed Type	
		7	Sweden	2013	92		Yearb I as at router			2012	94			
		8	Ireland		null		≥ 25-34 (%)[4]				null			
		9	Armenia	2010	93					null 2010	92			
		10	United Kingdom	2013	94					2013	92			
		11	Kazakhstan	2010	93					nul/ 2010	92			
		12	Poland	2013	83					2013	92			
		13	United States	2010	92					null 2013	91			
		14	Canada	2010	86					null 2010	91			
		15	Greece	2013	92					2013	91			
		16	Slovakia	2013	93					2013	90			
		17	Cyprus	2013	93					2013	90			
		18	Slovenia	2013	89					2013	90			
		19	Israel	2012	88			OK Cancel		null 2012	89			
		20	Russian Federation	2013	87					null 2013	88			
		21	Czech Republic	2013	90		2015	90.4		2013	88			
		22	Lithuania	2013	91		2015	90.9		2013	88			
		23	France	2013	83		2015	87.2		2013	88	~		
		24	<									>		
11 COLUMNIS 1	22 DOME Cal		in hard on ton 1000 -	7044									DESUEN DOWNLOW	DED AT SHO DM

Figure 13: Power Query Editor: Choosing Columns

Editing Rows

See the orange box in Figure 14. We removed the Year columns in the previous step, and we can see that the Power Query editor records this change. Choosing the gear icon to the right of each item, you can see the specific changes you made to the imported data. This is a massive improvement over manually editing data.

Next, we can remove rows that are irrelevant for our purposes. For this table, we can see that the variable names are repeated in the first row of the table. We can remove this row by clicking Remove Rows, then selecting Remove Top Rows in the red box. Remove the first row of this table by typing in 1 in the blue box, and click OK.

Image: state of the state	ery Editor Add Column View						- 0	× ^ •
Close & Load + Close & Load +	ttor Choose Remove Columns • Columns •	AL armove ows • Column •	Data Type: Text • Use First Row as Group By J 2 Replace Values	Headers • Append Queries •	Manage Parameters *	Data source settings	New Source •	
Close Query	Manage Columns Reduc	Remove top kows	Transform	Compine	Parameters	Data Sources	New Query	
Queries [1] < × √ fa	= Table.SelectColumns(#"Changed Type",	Remove Bottom Rows	tion age (%)[1]", "20-2	4 (%)[2]", "20-29 (%) [3]", "25-29 (%)	", "25-34 (%)[4]"})	~	Query Settings	×
III Table 0	A ⁸ C 3 to 5 Years above graduation ag	Remove Alternate Row:	S 🖓 C 20-29 (%) [3] 👻	A ⁸ _C 25−29 (%) ▼ A ⁸ _C 25−34 (%)[4	•			
1 Country	3 to 5 Years above graduation age	Remove Duplicates	20-29 (%)	25-29 25-34 (%) (%)		^	A PROPERTIES	
	(%)	Remove Blank Rows					Table U	
2 South Korea	99	Remove Errore	98	null	null		All Properties	
3 Georgia	96	X Remove chors	95	null	null		▲ APPLIED STEPS	
4 Japan	95		null nul	null	null		Source	\$
5 Croatia	95				× null		Navigation	*
6 Ukraine	95	Remove Top Rows			null		Changed Type	
7 Sweden	92	Conside how on the second					× Removed Other Columns	: - +>
8 ireland		specily now many rows to rem	ove from the top.					
9 Armenia	93	Number of rows			naii			
10 Onited Kingd	02				aul!			
12 Poland	83				11000			
13 United State	92			OK Case				
14 Canada	86			OK Calife				
15 Greece	92							
16 Slovakia	93	91.3	90	null 93				
17 Cyprus	93	94.3	90	null	null			
18 Slovenia	89	90.9	90	null 94				
19 Israel	88		null 89	null 91				
20 Russian Fede	ration 87		null 88	null 95				
21 Czech Repub	ic 90	90.4	88	null 94				
22 Lithuania	91	90.9	88	null 90				
23 France	83	87.2	88	null 87		~		
24 Belarus	83		nuil 88	null	null			

Figure 14: Power Query Editor

The other options included in either Keep Rows or Remove Rows may prove quite useful, and we encourage that readers try out these options:

- Keep / Remove Top Rows: Keep / Remove only the top N rows from
- Keep / Remove Bottom Rows: Keep / Remove only the bottom N rows from this table.
- Keep / Remove Range of Rows: Specify the number of rows to keep / remove starting at a specific row.
- Keep / Remove Duplicates: Keep / Remove rows containing duplicated values in the currently selected columns.
- Keep / Remove Errors: Keep / Remove only rows containing errors in the currently selected columns.

Loading Data to Excel

Once the table is edited to satisfaction, we can load it to Excel by clicking Close & Load in the red box. It is recommended that users check the Applied Steps in the blue box before loading the table to Excel.

File	• ∓ Tab Home	ole 0	- Power Query ranstorm Ad	Editor Id Colum	ו View									- 0	×
Close & Load •	Refresh Preview •		Properties Advanced Editor Manage •	Choos	e Remove s • Columns •	Keep Remove Rows • Rows •	Ą↓ Z↓	Split Grou Column • By	Data Type: Text • Use First Row	as Headers 🔻	Merge Queries •	Manage Parameters •	Data source settings	New Source •	
Close		Qu	lery	Mana	ge Columns	Reduce Rows	Sort		Iransform		Combine	Parameters	Data Sources	New Query	
Queries (1)		<	× √ fx -	Table.Ski	p(#"Removed Othe	r Columns",1)							~	Ouery Settings	~
Table 0			. Ar Country	- /	- 3 to 5 Years abov	e graduation age (%)[1]	Ψ A ^β r	20-24 (%)[2]	ABr 20-29 (%) [3]	▼ A ⁸ : 25-29 (%)	▼ / ⁸ r 25-34 (%)[4]	v		Query Settings	^
			1 South Korea		0	e Branannen albe (riv)[x]		10 11 (10)[1]	00			aul.			
			2 Georgia		6			10	y 95		null	null	^	Name	
			2 12020	-	e			10	v 55	sulf	oull	null		Table 0	
			4 Croatia		6		05	7	05	1011	null	null		All Properties	
			5 Ukraine	-	5			,	y 94		null	null			_
			6 Sweden	-	0		87	3	94		null 82	11000		▲ APPLIED STEPS	
			7 Ireland		-		null 92	7		Nall	null 91			Source	*
			8 Armenia		a			,	y 92		null	null		Navigation	*
			9 United Kingdom		4		85	7	92		null 85			Changed Type	
			10 Kazakhstan	-	4				y 92		null	null		Removed Other Columns	\$
			11 Poland		3		90	8	92		null 94				Ŕ
			12 United States		2				y 91		null 90				
			13 Canada	2	6			04	y 91		null 93				
			14 Greece		2		89	6	91		null 84				
			15 Slovakia		8		91	3	90		null 93				
			16 Cyprus	-	3		94	3	90		null	null			
			17 Slovenia	2	9		90	9	90		null 94				
			18 Israel	8	8				7 89		null 91				
			19 Russian Federation	8	7			nu	V 88		null 95				
			20 Czech Republic	9	0		90.	4	88		null 94				
			21 Lithuania	9	1		90.	9	88		null 90				
			22 France	8	3		87.	2	88		null 87				
			23 Belarus	8	3			nu	v 88		null	null			
			24 Austria	8	4		88.	7	88		null 90				
			25 Finland	8	5		86	8	87		null 90				
			26 Australia	8	5			nu	7 85		null 88		~		
6 COLUMNS 12			filing bared on top 1000	TOWS										PREVIEW DOWNI OAL	DED AT 5:40 PA

Figure 15: Loading Data to Excel

Figure 16 shows the data imported to Excel. The data will automatically be organized as a table as shown in the red box, and the default name will follow the object name we found in Figure 11.

File Home	nsert Page Layout Formulas I	Data Review View Automate	Help Table Desig	In Query								P	omments	년 Share 👻
Table Name: Table_0	Summarize with PivotTable Remove Duplicates Insert Convert to Range Tools	Export Refresh Solutions Solutions Export Refresh Solutions Solutions Export Table Data	Header Row Total Row Banded Rows	First Column Filter Button Last Column Banded Columns Table Style Options			Table Stvi	*						~
A1 v I	× √ fx Country													~
	A	В	D D	E F	G H	1.1	J	K L	M	N	O P			
1 Country	💌 3 to 5 Years abov	ve graduation age (%)[1] 🔽 20–24 (6)[2] 💌 20-29 (%)	[3] 💌 25-29 (%) 💌 25-34 (%)[4] 🕚								Queries & Connec	tions	~ ×
2 South Korea	99	• • • • • • • • •	98											
3 Georgia	96		95									Queries Connections		
4 Japan	95											1 00000		
5 Croatia	95	95.7	95									rigociy		
6 Ukraine	95		94									Table 0		G.
7 Sweden	92	87.3	94	82								121 rows loaded.		
8 Ireland		92.7		91										
9 Armenia	93		92											
10 United Kingdom	94	85.7	92	85										
11 Kazakhstan	93		92											
12 Poland	83	90.8	92	94										
13 United States	92		91	90										
14 Canada	86		91	93										
15 Greece	92	89.6	91	84										
16 Slovakia	93	91.3	90	93										
17 Cyprus	93	94.3	90											
18 Slovenia	89	90.9	90	94										
19 Israel	88		89	91										
20 Russian Endorst	ion 97		00	05										

Figure 16: Imported to Excel

Topic 3. Power Query: TXT, Splitting, Duplicating, & Grouping

We can also import text files via Power Query. Let us return to the text file we used previously, BUSI201-LEC20-txt. We may import a text file into Excel using the same Power Query framework by selecting From Text/CSV under the Data tab.

C B From Text/CSV E From Dicture x	D Queries & Connections
Get ☐ From Web © Recent Sources Ref Data ~ ☐ From Table/Range ☐ Existing Connections Al	Lefresh All ~ Device Connections Minor Connections Properties Stocks Currencies Geography

Figure 17: Loading Text / CSV to Excel

The window in Figure 18 should pop up when the text file is correctly selected. Since our text file is separated using commas, the delimiter is correctly set to Commas, and the preview in the blue box shows the correct layout for our table. Select Transform Data.

1252 Worter	European (Windows)			Data	a Type Detect
202: western	curopean (Windows) * Comma			- Bas	eu on Tirst 20
Code	Title	Year	Length	Rated	Rating
MDB-RANK-01	The Shawshank Redemption	1994	142	R	9.3
IMDB-RANK-02	The Godfather	1972	175	R	9.2
IMDB-RANK-03	The Dark Knight	2008	152	PG-13	9
MDB-RANK-04	The Godfather Part II	1974	202	R	9
MDB-RANK-05	12 Angry Men	1957	96	N/A	9
MDB-RANK-06	Schindler's List	1993	195	R	9
MDB-RANK-07	The Lord of the Rings: The Return of the King	2003	201	PG-13	9
IMDB-RANK-08	Pulp Fiction	1994	154	R	8.9
IMDB-RANK-09	The Lord of the Rings: The Fellowship of the Ring	2001	178	PG-13	8.8
MDB-RANK-10	The Good the Bad and the Ugly	1966	178	N/A	8.8
MDB-RANK-11	Forrest Gump	1994	142	PG-13	8.8
IMDB-RANK-12	Fight Club	1999	139	R	8.8
IMDB-RANK-13	The Lord of the Rings: The Two Towers	2002	179	PG-13	8.8
IMDB-RANK-14	Inception	2010	148	PG-13	8.8
IMDB-RANK-15	Star Wars: Episode V - The Empire Strikes Back	1980	124	PG	8.7
IMDB-RANK-16	The Matrix	1999	136	R	8.7
IMDB-RANK-17	Goodfellas	1990	145	R	8.7
IMDB-RANK-18	One Flew Over the Cuckoo's Nest	1975	143	R	8.7
IMDB-RANK-19	Se7en	1995	127	R	8.6
IMDB-RANK-20	It's a Wonderful Life	1946	130	PG	8.6
A					
Ihe data i	in the preview has been truncated due to size li	imits.			

Figure 18: Loading Text / CSV to Excel

Suppose that you want to create a column that splits the first column into many columns that has information on the ranking, which platform the rankings are based on, and the year when the rankings were taken. This can be acheived by splitting the Code column into many parts.

Splitting Columns

Select Split Column, and then choose By Delimiter, since the Code data is linked via short dashes. There are many different methods to split columns, and those methods may be useful depending on the type of data.

File	• ∓ BUSI2 Home	201-LEC20-txt - Pov Iransform Add	ver Query Editor Column View									- 0	×
Close & Load •	Refresh Preview •	Properties Advanced Editor Manage •	Choose Remove Columns * Columns *	Keep Remove Rows • Rows •	₽↓ Z↓ [Split Group	Data Type: Text 🕶	as Headers ▼ s	Merge Queries Merge Queries Append Queries Combine Files	Manage Parameters •	Data source settings	New Source Recent Sources Enter Data	
Close		Query	Manage Columns	Reduce Rows	Sort	By Delimite	r		Combine	Parameters	Data Sources	New Query	
Queries (1)	<	× √ fx = T	able.TransformColumnTypes(#"	Promoted Headers",	{{"Code",	By Number	of Characters	. Int64.Type}	<pre>}, {"Length", Int64.Type},</pre>	("Rated", type te	xt). Y	Quany Sattings	~
III BUSI20	1-LEC20-txt	III Ar Code	▼ A ⁸ r Title		¥ 122 ¥4	By Positions		r Alle Rated	× 1.2 Rating			Query settings	~
		1 IMDB-RANK-01	The Shawshank Redemption	0				12 R	ALL Houng	9.3			
		2 IMDB-RANK-02	The Godfather			By Lowerca	se to Uppercase	15 R		9.2	^	Name	
		3 IMDB-RANK-03	The Dark Knight			By Upperca	se to Lowercase	2 PG-13		9		BUSI201-LEC20-bit	
		4 IMDB-RANK-04	The Godfather Part II			D. Distant	les Dist	12 R		9		All Properties	
		5 IMDB-RANK-05	12 Angry Men			By Digit to I	Non-Digit	16 N/A		9		▲ APPLIED STEPS	
		6 IMDB-RANK-06	Schindler's List			By Non-Dig	it to Digit	15 R		9		Source	ŏ
		7 IMDB-RANK-07	The Lord of the Rings: The R	leturn of the King		2000		aJ1 PG-13		9		Promoted Headers	8
		8 IMDB-RANK-08	Pulp Fiction			1994		154 R		8.9		× Changed Type	
		9 IMDB-RANK-09	The Lord of the Rings: The F	ellowship of the Ring		2001		178 PG-13		8.8			
		10 IMDB-RANK-10	The Good the Bad and the U	Jgly		1966		178 N/A		8.8			
		11 IMDB-RANK-11	Forrest Gump			1994		142 PG-13		8.8			
		12 IMDB-RANK-12	Fight Club			1999		139 R		8.8			
		13 IMDB-RANK-13	The Lord of the Rings: The T	wo Towers		2002		179 PG-13		8.8			
		14 IMDB-RANK-14	Inception			2010		148 PG-13		8.8			
		15 IMDB-RANK-15	Star Wars: Episode V - The E	Empire Strikes Back		1980		124 PG		8.7			
		16 IMDB-RANK-16	The Matrix			1999		136 R		8.7			
		17 IMDB-RANK-17	Goodfellas			1990		145 R		8.7			
		18 IMDB-RANK-18	One Flew Over the Cuckoo's	s Nest		1975		143 R		8.7			
		19 IMDB-RANK-19	Se7en			1995		127 R		8.6			
		20 IMDB-RANK-20	It's a Wonderful Life			1946		130 PG		8.6			
		21 METACRITIC-RANK-01	Citizen Kane			1941		119 PG		100			
		22 METACRITIC-RANK-02	The Godfather			1972		175 R		100			
		23 METACRITIC-RANK-03	Rear Window			1954		112 PG		100			
		24 METACRITIC-RANK-04	Casablanca			1942		102 PG		100			
		25 METACRITIC-RANK-05	Boyhood			2014		165 R		100	~		
		26 METACRITIC-RANK-06	Three Colors: Red			1994		99 R		100			
6 COLUMNS, 40	ROWS Column p	rofiling based on top 1000 rov										PREVIEW DOWNLO	ADED AT 9:28 PM

Figure 19: Splitting Columns in Power Query

We can tell Excel which delimiter will be used to split the column in the red box in Figure 20. Set up the options as shown in the red box and blue box to split the column Code. Choosing any of the other options in the blue box will allow the user to split the column in various ways.

File	Home	Iransform Ac	ld Column View		~ ?
Close & F Load • Pr	Refresh review •	Properties Advanced Editor Manage 🗸	Choose Remove Columns * Columns *	Image: Split Split Group Data Type: Text * Image: Split Image: Split Image: Split Manage: Split Manage: Split Manage: Split Image: Split Image	New Source ▼ Recent Sources ▼ Enter Data
Close	0	Query	Manage Columns	Parameters Data Sources	New Query
Queries (1)	C20-txt	→ √ fx •	Table.TransformColumnTypes(#	Split Column by Delimiter , ("Mated", type text), v	Query Settings ×
		2 IMDB-RANK-02	The Godfather	Select or enter delimiter 9.2	Name
		3 IMDB-RANK-03	The Dark Knight	9	BOSI201+LEC20+Dit
		4 IMDB-RANK-04	The Godfather Part II		All Properties
		5 IMDB-RANK-05	12 Angry Men	Split at 9	A APPLIED STEPS
		6 IMDB-RANK-06	Schindler's List	C Left-most delimiter 9	Source (5
		7 IMDB-RANK-07	The Lord of the Rings: The	Right-most delimiter g	Bromoted Handerr Ö
		8 IMDB-RANK-08	Pulp Fiction	Each occurrence of the delimiter	X Changed Type
		9 IMDB-RANK-09	The Lord of the Rings: The	8.8	energee type
		10 IMDB-RANK-10	The Good the Bad and the	Advanced options 8.8	
		11 IMDB-RANK-11	Forrest Gump	Ounte Character 8.8	
		12 IMDB-RANK-12	Fight Club	* * 8.8	
		13 IMDB-RANK-13	The Lord of the Rings: The	8.8	
		14 IMDB-RANK-14	Inception	Split using special characters 8.8	
		15 IMDB-RANK-15	Star Wars: Episode V - The	Insert special character v 8.7	
		16 IMDB-RANK-16	The Matrix	8.7	
		17 IMDB-RANK-17	Goodfellas	OK Cancel 8.7	
		18 IMDB-RANK-18	One Flew Over the Cuckoo	8.7	
		19 IMDB-RANK-19	Se7en	1.000 1.00 Mar 14 8.6	

Figure 20: By Delimiter Options

Clicking OK, the column Code will be split into three parts as shown in the <u>red box</u> in Figure 21. The original column has been split by each occurrence of -, creating columns named Code . 1, Code . 2, and Code . 3. You may double click the header containing the names of the columns to rename the columns.

File	+ ∓ BUS Home	l201- Irai	LEC20-txt - Po nstorm Ad	wer Que	ry Editor View										— 0	× ~ •
Close & Load • Close	Refresh Preview •	Pro Ad Ma	operties vanced Editor mage •	Choose Column Manae	Remove Columns • ge Columns	Kee Row	p Remove s Rows V	Q↓ Z↓ Sort	Split G Column •	Data Type: Text • Tube: Text • Use First Row a: Tube: Tube: Text • Use First Row a: Tube: Tube: T	Headers 🔻	Merge Queries • Append Queries • Combine Files	Manage Parameters • Parameters	Data source settings Data Sources	New Source • Recent Sources • Enter Data	
Queries (1)	<		√ fx =	Table.Tran	sformColumnType	(#"Spli	t Column by De	limiter".	{{"Code.1", t	<pre>vpe text}, {"Code.2", type to</pre>	xt), {"Code.3"	<pre>, Int64.Type}})</pre>		~	Quan Cattings	
III BUSI20	11-LEC20-txt	-	Ally Code 1	× 4	- Code 2		2. Code 2		Ale Title		12. Year	v 12. Length	T Mr. Pat	ad a	Query settings	^
		1004	HC COULT		C COULLE		1-3 COULS		ac nuc		1-3 100	- 1-3 tengen	- m; nu	eu		
		1	IMDB	R	ANK				The Shawshank	Redemption		1994	142 R	~	Name	
		2	IMDB	R	ANK				The Godfather			1972	175 R		BUSI201-LEC20-txt	
		3	IMDB	R	ANK			-	The Dark Knight			2008	152 PG-13		All Properties	
		4	IMDB	R	ANK				The Godfather P	art II		1974	202 K	_		
		2	INIDO	K	ANK			-	12 Angry Men			1957	90 N/A		▲ APPLIED STEPS	
		0	IMDB	R	ANK				Schindler's List	North Participation of the Mine		1993	195 K	_	Source	\$
		-	INIDO	K	ANK				The Lord of the s	ongs: The Return of the King		2005	201 PG-15		Promoted Headers	*
		-	INIDO		SINK .				Pulp Piction	New The Collected is of the Dire		1994	134 N	_	Changed Type	
		9	INIDO	K					The Condition Re	kings: The Fellowship of the King		2001	178 1/4		Split Column by Delimiter	4
		10	IMDB						Formert Cump	d and the ogy		1900	1/0 N/A		➤ Changed Type1	
		10	INIDO						finest dump			1000	142 PG-15			
		12	INIDO					**	The Lord of the l	lings: The Two Toward		2002	139 N			
		10	INIDO					1.	The bord of the s	angs, the two towers		2002	1/3 PG-13			
		10	IMDR						Star Marri Enico	do V. The Empire Striker Back		1020	124 00			
		16	IMDB					1	The Matrix	de v - me Empire Scrikes back		1990	124 PG	_		
		17	IMDR	0				1	Condialias			1000	145 D			
		10	IMDB	0	NNK .			11	One Eleve Over t	ha Curkoo'r Nert		1075	142 P			
		10	IMDB	0	INK			10	Se7en	ne cachoo s mest		1995	127 P			
		20	IMDB	R	INK			21	It's a Wonderful	life		1945	130 PG			
		20	METACRITIC	0	NK .			-	Citizen Kane	101 W		10/1	110 PG			
		21	METACRITIC	R	INK				The Godfather			1972	175 R			
		23	METACRITIC	R	ANK				Rear Window			1954	112 PG			
		24	METACRITIC	R	INK				Casablanca			1942	102 PG			
		25	METACRITIC	8	ANK				Boyhood			2014	165 B	~		
		20		15		_							200 1			

Figure 21: Splitting Code

Creating Duplicate Queries

We can create duplicate queries by right clicking the original query in the <u>red box</u> in Figure 22, and clicking <u>Duplicate</u>. We will later be using these duplicates to generate new variables, and merge data.

×	• - BU:	SI201	LEC20-txt - Powe	er Query Edi	tor								- 0	×
Close & Load • Close	Refresh Preview •	Pr Ac M Que	operties dvanced Editor anage • Co ry	Choose Re olumns • Col Manage Colu	emove Ke umns • Rov	ep Remove vs • Rows • duce Rows	AJ AJ Sort	Split Group Column - By Transform	i Headers 🔻	Append Queries - Append Queries - Combine Files Combine	Manage Parameters • Parameters	Data source settings Data Sources	New Source • Recent Sources • Enter Data New Query	
Oueries (11		< 7	<	le.TransformCo	lumnTypes(#"Sp]	it Column by De	limiter"	,{{"Code.1", type text}, {"Code.2", type te	xt}, {"Code.3"	, Int64.Type}})		~	Ouerv Settings	×
BUSI20	11-LEC20-txt		Mc Code.1	▼ A ^{ll} c Code.2		1 ² 3 Code.3		All _C Title	1 ² 3 Year	▼ 1 ² 3 Length	✓ A ⁰ c Rate	ed 🔽	()g-	
								1 The Shawrhank Redemotion		1004	142 P		▲ PROPERTIES	
		h	Сору					2 The Godfather		1972	175 P	^	Name	
		1 B	Paste					3 The Dark Knight		2008	152 PG-13		BUSI201-LEC20-txt	
								4 The Godfather Part II		1974	202 R		All Properties	
		\times	Delete					5 12 Angry Men		1957	96 N/A			
		<u>آت</u>	Rename					6 Schindler's List		1993	195 R		A AFFLIED STEPS	
		-	Renderic					7 The Lord of the Rings: The Return of the King		2003	201 PG-13		Source	8
		6	Duplicate					8 Pulp Fiction		1994	154 R		Promoted Headers	8
		9	Reference					9 The Lord of the Rings: The Fellowship of the Ring		2001	178 PG-13		Changed Type Solit Column by Dolimitor	
		0	nererere				1	0 The Good the Bad and the Ugly		1966	178 N/A		X Changed Type1	
			Move To Group	· · ·			1	1 Forrest Gump		1994	142 PG-13		- Consinged Type I	
			Mayalla				1	2 Fight Club		1999	<i>139</i> R			
			wove op				1	3 The Lord of the Rings: The Two Towers		2002	179 PG-13			
			Move Down				1	4 Inception		2010	148 PG-13			
			Croate Eurotion				1	5 Star Wars: Episode V - The Empire Strikes Back		1980	124 PG			
			create runction.				1	6 The Matrix		1999	<i>136</i> R			
			Convert To Para	meter			1	7 Goodfellas		1990	145 R			
		B	Advanced Edito				1	8 One Flew Over the Cuckoo's Nest		1975	143 R			
		LE	Advanced Edito	"			1	9 Se7en		1995	127 R			
		E,	Properties				2	0 It's a Wonderful Life		1946	130 PG			
		21	METACRITIC	KANK				1 Citizen Kane		1941	119 PG			
		22	METACRITIC	RANK				2 The Godfather		1972	175 R			
		23	METACRITIC	RANK				3 Rear Window		1954	112 PG			
		24	METACRITIC	RANK				4 Casabianca		1942	102 PG	~		
		25	METACRITIC	RANK				5 boynood		2014	100 R			
		26	× –				_					>	1	

Figure 22: Creating Duplicates

Grouping

We can use the grouping tool to generate new variables based on this data. Suppose we wanted to know how many movies are in the top movies by its ratings; R, PG, PG-13, etc. Click Group By, and setting up the options as shown in the blue box in Figure 23.³

File		201-LEC20-txt - Po Transform Ad	ower Query Editor Ia Column View							- 0	× ^ ♥
Close & Load •	Refresh Preview •	Properties Advanced Editor Manage T	Choose Remove Columns * Columns *	Keep Remove Rows * Rows *	Ž↓ Ž↓ Split Column	Group By By Data Type: Text *	 Merge Queries * Append Queries * Combine Files 	Manage Parameters •	Data source settings	New Source Recent Sources Enter Data	
Close		Query	Manage Columns	Reduce Rows	Sort	Transform	Combine	Parameters	Data Sources	New Query	
Queries [1]	<	× √ /x -	Table.TransformColumnTypes(#"	Solit Column by Deli	miter".{{"Code.1	. type text), ("Code.2", type text), ("Cod	e.3". Int64.Type}})		~	Ourse Catting and	
🔥 BUSI201-	LEC20-brt	III Alle Code 1	w the Code 2	x 12x Code 2	w Alle Title	- 1 ² , Var	w 12a Longth	v dla Para	4	Query settings	~
		1 11408	DANK	13 0000	inc nuc	13 100	×	143.0		▲ PROPERTIES	
		2 IMDB	PANK	Course Day				175 P	^	Name	
		3 IMDB	RANK	Group By				152 PG-13		BUSI201-LEC20-txt	
		4 IMDB	RANK	Specify the column	n to group by and t	he desired output.		202 R	_	All Properties	
		5 IMDB	RANK	Basic O Advan	iced			96 N/A			
		6 IMDB	RANK					195 R		A APPLIED STEPS	
		7 IMDB	RANK	0.1	-			201 PG-13		Source	8
		8 IMDB	RANK	Rated				154 R		Promoted Headers	~
		9 IMDB	RANK	New column name	0	peration Column		178 PG-13		Split Column by Delimiter	8
		10 IMDB	RANK	Count	(Count Rows *	v	178 N/A	_	× Changed Type1	
		11 IMDB	RANK					142 PG-13		5 51	
		12 IMDB	RANK					139 R			
		13 IMDB	RANK				OK Cancel	179 PG-13	_		
		14 IMDB	RANK					148 PG-13			
		15 IMDB	RANK					124 PG			
		16 IMDB	RANK		16 The Matrix		1999	136 R			
		17 IMDB	RANK		17 Goodfellas		1990	145 R			
		18 IMDB	RANK		18 One Flew C	Over the Cuckoo's Nest	1975	143 R			
		19 IMDB	RANK		19 Se7en		1995	127 R			
		20 IMDB	RANK		20 It's a Wond	lerful Life	1946	130 PG			
		21 METACRITIC	RANK		1 Citizen Kan	e	1941	119 PG			
		22 METACRITIC	RANK		2 The Godfat	ther	1972	175 R			
		23 METACRITIC	RANK		3 Rear Winds	DW	1954	112 PG			
		24 METACRITIC	RANK		4 Casablanca		1942	102 PG	~		
		25 METACRITIC	RANK		5 Boyhood		2014	165 R			
		26 🔇							>		

Figure 23: Grouping by Rated

Figure 24 shows us how the data will be transformed following the grouping described above. We will later see how we can merge query tables to consolidate multiple data sources.

File	• - Cou Home	nt by Ratings - Pov Iranstorm Ad	ver Query I Id Column	ditor View									- 0	×
Close & Load •	Refresh Preview •	Properties Advanced Editor Manage -	Choose Columns •	Remove Columns *	Keep Rem Rows • Row	× Z↓ Z↓ Z↓ s ▼	Split Column •	Group By	Data Type: Whole Number ▼ Use First Row as Headers ▼ ¹ → ₂ Replace Values	Merge Queries Append Queries Combine Files	Manage Parameters •	Data source settings	New Source Recent Sources Enter Data	
Close		Query	Manage	Columns	Reduce Row	s Sort			Transform	Combine	Parameters	Data Sources	New Query	
Courtes (2)	-LEC20-bxt yy Ratings	X V K III. Mc Rased I I R 2 PG-13 3 N/A H G 5 G G G	Jable.Group(#"Changed Type Count	1". ("Rated"), { 	("Count", e	the trade.Row	Count(_),	Int64.Type)))			~	Query Settings 9 POPERTIS Name Court by Rangs All Poperties A PPUED STEPS Source Headers Charged Type Splt Column by Delimiter Charged Type 1 X Grouped Roos	×

Figure 24: Grouped by Rated

³At this point, we ignore the issue of duplicates. For instance, The Godfather is included in all three lists.

Topic 4. Power Query: Refreshing Data

You might still believe that manually importing data isn't too troublesome, considering the need to learn another tool. However, data imported using Power Query offers a crucial advantage over manual imports – it enables us to refresh the tables in Excel when the source data changes. Import the table using Power Query from **Topic 3** to obtain the two worksheets depicted below.



Suppose now that the rating for the top-rated movie on IMDB, "The Shawshank Redemption," has been updated to 9.5. Additionally, let's assume that "The Dark Knight" has had its PG-13 rating updated to an R rating. Open the file BUSI201-LEC20-txt, update the ratings in the source file, and then save the file.

	BUSI201-LEC20-txt	٠	+
File	Edit View		
Code IMDE IMDE IMDE IMDE IMDE	e,Title,Year,Length,Rated, 3-RANK-01,The Shawshank Re 3-RANK-02,The Godfather,19 3-RANK-03,The Dark Knight, 3-RANK-04,The Godfather Pa 3-RANK-05,12 Angry Men,195	Rati demp 72,1 2008 art I 57,96	ng tion,1994,142,R <mark>.9.5</mark> 75,R,9.2 ,152 <mark>R</mark> 9 I,1974,202,R,9 ,N/A,9

Figure 25: Updating Source Data



Figure 26: Refreshing Tables

Return to the Excel worksheet BUSI201-LEC20-txt containing the main source, click the tables generated using Power Query to make the Query tab available. Then, click Refresh to update the table.

	А	В	С	D	E	F	G	Н
1	Code.1 🗾	Title 🔽	Year 💌	Length 💌	Rated 💌	Rating 💌		
2	IMDB	The Shawshank Redemption	1994	142	R	9.5		
3	IMDB	The Godfather	1972	175	R	9.2		
4	IMDB	The Dark Knight	2008	152	R	9		
5	IMDB	The Godfather Part II	1974	202	R	9		
6	IMDB	12 Angry Men	1957	96	N/A	9		
7	IMDB	Schindler's List	1993	195	R	9		

Figure 27: First Table Refreshed

See Figure 27 to observe the refreshed table. Follow the same workflow to update the second table, where we grouped the data. Once you refresh the second table in the worksheet BUSI201-LEC20-txt(2), you can see that the grouped table displays the updated distribution as shown in Figure 28.

	А	В	С	D	E
1	Rated 💌	Count 💌			
2	R	16			
3	N/A	7			
4	PG-13	5			
5	PG	9			
6	G	3			
7					

Figure 28: Second Table Refreshed

While we cannot demonstrate here, if information on a webpage is updated, you may update the contents of your Excel spreadsheet by clicking refresh.