# SUMMARIES AND VISUALIZATIONS

The notion of income inequality has received considerable attention in recent years. The gap between those that have a lot of money and those that don't continues to increase in the United States.

**Income inequality** refers to the extent to which **income** is distributed in an uneven manner among a population. In the United States, **income inequality**, or the **gap** between the rich and everyone else, has been growing markedly, by every major statistical measure, for some 30 years.



The United States Census Bureau releases aggregated data on a regular basis for states and counties in the United States. Consider the following graph that shows Median Household Income (for years 2009-2013) by State. We see that NJ has the highest household income and Mississippi has the lowest at about \$35,000.



Forbes has recently published several articles centered on whether or not a college education is worth it. For many, a college education is necessary to find their first job. However, other graduates are over qualified for their job and are hampered by a substantial amount of student loan debt. The following table suggest that when a large percentage of county residents have a bachelor's degree or more, the typical household income is over \$50,000. When this percentage is low, the typical income level drops to \$38,000 per household.

Average
Income
\$54,045
\$44,689
\$38,074
\$45,937

The data from this handout is provided by the United States Census Bureau. This data will need to be imported into Excel. The information needed for our analysis is not contained in a single file, but two different files. The auxiliary information contained in the FIPS Codes dataset will need to be merged with the dataset before summaries and visualization can be constructed.

**Procedural Steps** 

- 1. Download data and FIP Codes files from the US QuickFacts website
- 2. Merge the FIPS Code information with the data
- 3. Create new variables for County and State
- 4. Construct various summaries and visualizations in Excel

Data Technologies

- 1. Import file into Excel
- 2. Summaries and Visualizations through PivotTables

Data Source								
Address	http://quickfacts.census.gov/qfd/download_data.html							
Description	US QuickFacts							
	The US QuickFacts dataset contains aggregate information for several variables at the County and State level. The variable Median Household Income (2009-2013) will be the focus of this investigation.							
	Download Data Dictionary							

First, let us consider the DataSet.txt file from their website. This data may not appear to have much structure. This data has been provided to us in a format known as a comma delimited or comma separated value format, i.e. csv format.

📃 Qu	ickFa	cts_Data ·	- Notepa	d							
File	Edit	Format	View	Help							
fips 00000 01000 01000 01000 01000	,PST 0,31 0,48 1,55 3,20 5,26	045214 885705 49377, 395,55 0111,1 887,26	,PST0 6,316 48339 136,5 95443 978,2	45213 497533 96,478 4571,3 ,18220 7457,-	PST04 ,3087 30127 ,5,1. 55,9.8 -2.1,-	0210 5810 1.4,1 0,545 3,7.2 -1.7,2	PST12 ,3.3, .1,47 571,6. ,18226 27457,	0214 2.5, 7973 1,25 5,5. 5.8,	,PST1 30874 6,6.1 .4,13 7,22. 21.1	L2021 45538 L,23. 3.5,5 4,18 ,15.9	L3,P 3,6. 0,1 51.5 3.1, 9,46

Excel has the ability to directly import this type of file. This process is started by selecting Data > From Text.



A Text Import Wizard window will show up. Proceed through this wizard by specifying the following in each step.

In Step 1 of 3, select Delimited

Specify that the Delimiter is Comma and click Finish

Text Import Wizard - Step 1 of 3	? 💌	Text Im	oort Wizard - St	tep 2 of 3					-	? <b>x</b>
The Text Wizard has determined that your data is Delimited. If this is correct, choose Next, or choose the data type that best describes your data. Original data type		This scr preview Delimi	en lets you set below. ers	the delimiter:	s your data co	ntains. You ca	an see how yo	our text is affe	ted in th	e
Choose the file type that best describes your data:      © <u>Delimited</u> - Characters such as commas or tabs separate each field.      Fixed width - Fields are aligned in columns with spaces between each field.		Ia Se Co Se Se	nicolon mma ace	Treat cons	ecutive delimi	ters as one				
Start import at row: 1 File origin: 437 : OEM United States	•		her:							
My data has headers.		Data <u>r</u>	review							
Preview of file W:\workshops\uscots2015\SummaryViz\QuickFacts_Data.txt.										
1       Fips, DST045214, DST045213, DST040210, DST120214, DST120213, DOP010210,         2       D0000, 318857056, 316497531, 308758105, 3.3, 2.5, 308745588, 6.3, 2.3, 3, 14         3       D1000, 4849377, 4833996, 4780127, 1.4, 1.1, 4779736, 6.1, 2.3, 0, 14.9, 51.5, 410101, 58395, 55136, 54571, 1.5, 1.0, 54571, 1.6, 1.2, 54, 13.5, 51.5, 78.1, 18.5         5       D1003, 200111, 195443, 182265, 9.8, 7.2, 182265, 5.7, 22.4, 18.1, 51.2, 87.3	AGE13521 A .1,50.8, 69.8,26. 4,0.5,1. ,9.5,0.7	fips 0000 0100 0100 0100	PST045214 318857056 4849377 55395 3 200111	PST045213 316497531 4833996 55136 195443	PST040210 308758105 4780127 54571 182265	PST120214 3.3 1.4 1.5 9.8	PST120213 2.5 1.1 1.0 7.2	POP010210 308745538 4779736 54571 182265	AGE1352 6.3 6.1 6.1 5.7	*
Cancel < Back <u>N</u> ext >	<u> </u>					Cancel	< <u>B</u> ack	<u>N</u> ext >	<u><u> </u></u>	nish

The last step is to tell Excel where you'd like the dataset to be placed. You can specify the cell location in an existing workshop or select New Workshet.

Import Data							
Select how you want to view this data in your workbook.							
📰 🍥 Table							
📝 🔘 PivotTable Report							
📑 🔘 PivotChart							
Only Create Connection							
Where do you want to put the data?							
Existing worksheet:							
=SAS1							
New worksheet							
Add this data to the Data Model							
Properties OK Cancel							

	А	Б	L	υ	E	F	6	н	1	J
1	fips	PST045214	PST045213	PST040210	PST120214	PST120213	POP010210	AGE135213	AGE295213	AGE775213
2	0	318857056	316497531	308758105	3.3	2.5	308745538	6.3	23.3	14.:
3	1000	4849377	4833996	4780127	1.4	1.1	4779736	6.1	23	14.9
4	1001	55395	55136	54571	1.5	1	54571	6.1	25.4	13.
5	1003	200111	195443	182265	9.8	7.2	182265	5.7	22.4	18.
6	1005	26887	26978	27457	-2.1	-1.7	27457	5.8	21.1	15.9
7	1007	22506	22504	22919	-1.8	-1.8	22915	5.3	21.3	14.:
8	1009	57719	57720	57322	0.7	0.7	57322	6.1	23.8	16.4
0	1011	10764	10605	10015	1 /	20	1001/	6.2	21	14 (

Click OK and the contents of the DataSet.txt should be successfully imported into Excel. The following snippet is given for reference.

Unfortunately the only reference to county is through the Federal Information Processing Standard (FIPS) code provided in Column A. A FIPS codes is a five-digit code which uniquely identifies counties and county equivalents in the United States. States are given FIPS codes as well.

The FIPS\_CountyName.txt file contains the information necessary to relate a FIPS code to a county or state name. Before the file is imported, we must insert a column for the contents of this file.





An empty column should be provided

	А	В	С	D
1		fips	PST045214	PST045
2		0	318857056	31649
3		1000	4849377	483
4		1001	55395	5
5		1003	200111	19
6		1005	26887	20
7		1007	22506	2:
8		1009	57719	5

The FIPS\_CountyName.txt file format is a different format than file containing the data. In particular, the first 5 digits contain the FIPS code. A comma is used to separate the County Name from the State Name.

QuickFacts_FIPS - Notepad
File Edit Format View Help
00000 UNITED STATES
01000 ALABAMA
01001 Autauga County, AL
01003 Baldwin County, AL
01005 Barbour County, AL
01007 Bibb County, AL
01009 Blount County, AL
01011 Bullock County, AL
01013 Butler County, AL
01015 Calhoun County, AL

In Step 1 of the Text Import Wizard, Fixed width should be selected.

Text Import Wizard - Step	p 1 of 3	Į	? 💌
The Text Wizard has dete	rmined that your data is De	elimited.	
If this is correct, choose	Specify	that best describes your data.	
Choose the file type th	Fixed width		
Fixed width	for FIP Codes	tabs separate each field. with spaces between each field.	
Start import at row: 1	File <u>o</u> rigin:	437 : OEM United States	•
My data has headers.			
Preview of file W:\work	shops\uscots2015\Summar	yViz\QuickFacts_FIPS.txt.	
1 00000 UNITED ST 2 01000 ALABAMA 3 01001 Autauga C	ATES County, AL		Â
4 01003 Baldwin C 5 01005 Barbour C	ounty, AL ounty, AL		-
*			+
	Ca	ancel < Back Next >	<u>F</u> inish

In Step 2, specify that the first five column of each row should be separated from the remaining information. Click Next.

Text Imp	ort Wizard	- Step 2 of	3					?	X
This scre Lines wit	en lets you th arrows si	set field wi gnify a colu	dths (colum Imn break.	ın breaks).					
To Cf To Di To M Data <u>p</u> r	REATE a brea ELETE a brea IOVE a brea review 10	ak line, clici ik line, dou k line, click 20	k at the des ble click or and drag it	ired position. the line.  30	40	50	60	70	
00000 01000 01001 01003 01005	UNITED ALABAMA Autauga Baldwin Barbour	County, County, County,	AL AL AL						
				Cance		< <u>B</u> ack	Next >	<u> </u>	sh

In the Import Window, specify you want the information placed in cell A1.

The information from the FIPS\_CountyName.txt file should now be placed into the first two columns.

Α	В	С	D	E
0	UNITED STATES	fips	PST045214	PST045213
1000	ALABAMA	0	318857056	316497531
1001	Autauga County, AL	1000	4849377	4833996
1003	Baldwin County, AL	1001	55395	55136
1005	Barbour County, AL	1003	200111	195443
1007	Bibb County, AL	1005	26887	26978
1009	Blount County, AL	1007	22506	22504
	A 00 1000 1001 1003 1005 1007 1009	AB0UNITED STATES1000ALABAMA1001Autauga County, AL1003Baldwin County, AL1005Barbour County, AL1007Bibb County, AL1009Blount County, AL	ABC0UNITED STATESfips1000ALABAMA01001Autauga County, AL10001003Baldwin County, AL10011005Barbour County, AL10031007Bibb County, AL10051009Blount County, AL1007	A         B         C         D           0         UNITED STATES         fips         PST045214           1000         ALABAMA         0         318857056           1001         Autauga County, AL         1000         4849377           1003         Baldwin County, AL         1001         55395           1005         Barbour County, AL         1003         200111           1007         Bibb County, AL         1005         26887           1009         Blount County, AL         1007         22506

Unfortunately, the FIPS\_CountyName.txt file did not contain a header row like the DataSet.txt file did. Thus, all rows in columns A and B will need to be shifted down one row.

	Α	В	С	D	E
1	0	UNITED STATES	fips	PST045214	PST04521
2	1000			318857056	31649753
3	1001	Autauga County, AL	1000	4849377	483399
4	: 3	Palahuin Caustry Al	1001	55395	5513
5	5	Shift Columns A & B	1003	200111	19544
6	1007	down one row	1005	26887	2697
7	1009	BIOUNT COUNTY, AL	1007	22506	2250

<u>Note</u>: You could import the FIPS\_CountyName.txt file a second time and specify the locations in the Import Window to be A2 instead of A1.

To add a row for only Columns A and B, highlight cells A1 and B1, right click, and select Insert. Specify Shift cells down to insert a row at the top of columns A and B.

Insert 💦 💌
Insert
Shift cells right
Shift cells <u>d</u> own
Entire row
Entire <u>c</u> olumn
OK Cancel

Specify variable names for these new columns. FIPS2 and Location were used in my dataset.

	Α	В	C	D	E	F	G
1	FIPS2	Location	fips	PST045214	PST045213	PST040210	PST120214
2	0	UNITED STATES	0	318857056	316497531	308758105	3.3
3	1000	ALABAMA	1000	4849377	4833996	4780127	1.4
4	1001	Autauga County, AL	1001	55395	55136	54571	1.5
5	1003	Baldwin County, AL	1003	200111	195443	182265	9.8
6	1005	Parhour County Al	1005	26007	26070	27/157	2.1

# **Getting Summaries Using PivotTables**

This dataset contains several variable or fields. The names of these fields are abbreviated substantially as is often the case. A **data dictionary** is often provided with complex datasets. This dictionary contains detailed information about each variable. The dictionary for this data is provided in the DataDict.txt file. A review of this file informs us that INC110213 contains the Median Household Income, which is the variable of interest here.

HSG495213 Median value of owner-occupied housing units, 2009-2013 HSD410213 Households, 2009-2013 HSD310213 Persons per household, 2009-2013 INC910213 Per capita money income in past 12 months (2013 dollars), 2009-2013 INC110213 Median household income, 2009-2013 PVY020213 Persons below poverty level, percent, 2009-2013 BZ2010213 Private ponfarm establishments 2013

Select Insert > PivotTable (Data > PivotTable on a MAC). Construct a pivot table using the following structure.



It appears the average median household income value is a little over \$46,000. The count identifies that there were 3,195 observations used when computing this average.



However, the value for the United States, the other States, and Washington DC have been incorrectly included in this average.

If summaries are to be computed only on county level data, then a new variable should be created to identify whether or not the information provide in that row from a country.

FIPS Code	Location	Description			
0	UNITED STATES	FIPS code for US is 0			
1000	ALABAMA	FIPS code for AL			
1001	Autauga County, AL	1 <sup>st</sup> county in AL			
1003	Baldwin County, AL	2 <sup>nd</sup> count in AL			
1099	Monroe County, AL				
skinned		1100 is skipped as hundreds denotes a state			
зкіррей		for most instances (see note below)			
1101	Montgomery County, AL				
1133	Winston County, AL				
2000	ALASKA	FIPS code for AK			
2013	Aleutians East Borough,	1 <sup>st</sup> county equivalent in AK			
AK					
Note: Four exceptions include: 2100 Haines Borough, AK; 51600 Fairfax City, VA; 51600					
Newport News City, VA; 51800 Suffolk City, VA					

# Creating a New Variable for County

The following will be used to create a new variable to identify whether or not the row contains information for a county.

Insert a new column after Location. Type the following formula into cell C2.

Modular arithmetic is used because it is know that for all but four counties this function will produce a nonzero value. After this formula is entered, place the cursor back into cell C2 and double click on the lower-right corner. This will autofill the formula for all remaining rows.

In cell D2 enter the following formula which simply checks whether or not the value in Cell C2 is 0. If this value is 0, then we know that for all but four FIPS codes this value will not be a county. "No" is returned when the condition being checked is TRUE and "Yes" when the condition is False.

```
С
      A
                     В
                                                            D
1 FIPS2 🔻
                  Location
                               Ŧ
                                     Mod MATH
                                                          County
                                                 Ŧ
                                                                      Ŧ
                                 =MOD(A2,100)
2
         0 UNITED STATES
                                                   =IF(C2=0,"No","Yes")
3
      1000 ALABAMA
4
      1001 Autauga County, AL
5
      1003 Baldwin County, AL
6
      1005 Parhour County, AL
```

```
Cell D2: =IF(C2=0,"No","Yes")
```

Before proceeding, the four exceptions to the rule that "hundreds denotes state FIPS code" should be fixed. This can be done by applying a filter to the FIPS code column and select values 2100, 51600, 51700, and 51800. Simply change the value in Column C to a non-zero value, e.g. change them to 1.

	Α	В	С		D	
1	FIPS2 🖵 Location 💌		Mod MATH	•	County	-
80	2100	Haines Borough AK		0 Io		
2975	51600	Change to nonz	ero value	o lo		
2988	51700	Eventions not	d above (	o lo		
2999	51800	Exceptions note		o lo		
	А	В	C	L	D	
1	FIPS2 🖵	Location 💌	Mod MATH		County	-
80	2100	Haines Borough, AK	1	Yes		
2975	51600	Fairfax city, VA	1	Yes		
2988	51700	Newport News city, VA	1	Yes		
2999	51800	Suffolk city, VA	1	Yes		

### <u>Questions</u>

1. Use PivotTables to verify that the average median household income across all counties in the United States is about \$46,000.

PivotTable structure			0	utcomes	
Drag fields between areas below:				Average	
▼ FILTERS			County ? 💌	Income	Count
	∑ Values ▼		No	53520.96	52
			Yes	45937.12	3143
■ ROWS	$\Sigma$ VALUES		Grand Total	46060.55	3195
County •	Average of INC110213   Count of INC110213_2				

2. There is a Wikipedia page that lists information regarding all counties and county equivalents in the United State. The following text is from this Wikepedia page. Note: The table provided on the Wikipedia page is missing one entry. FIPS Code 51515 Bedford City, VA is missing. Please don't ask how I discovered this!

This is a complete list of the 3,143 counties and county equivalents of the United States of America as of July 1, 2013. For more detailed information, see the individual state lists shown below.

Source: http://en.wikipedia.org/wiki/List of United States counties and county equivalents

Does it appear that the average computed above is using the correct number of rows? Discuss.

3. The County = No has 52 entries; however, there are only 50 states. Determine why there are 52 rows labeled as "No" by our procedure.

# Creating a New Variable for State

A process similar to labeling Counties can be used to create a new variable for State abbreviations, i.e. AL, AK, etc. This abbreviation is present in Column B when the row consists of a county. In fact, this abbreviation is always the last two digits in the string. The =RIGHT() function will be used to pull off the last two digits, when appropriate, from column B.

- IF statement used to check whether or not column C contains a 0
- IF cell C2 is zero, then an empty string is returned, i.e. nothing is returned
- IF cell C2 is non-zero, then use the =RIGHT() function to pull off the last two digits

The following snippet is provided for reference.

	А	В	С	D	E
1	FIPS2	Location	Mod MATH	County	State
2	0	UNITED STATES	0	No	=IF(C2=0,"",RIGHT(B2,2))
3	1000	ALABAMA	0	No	
4	1001	Autauga County, AL	1	Yes	AL
5	1003	Baldwin County, AL	3	Yes	AL
6	1005	Barbour County. AL	5	Yes	AL

Verify the content is correct for the new variable State for several rows.

	Α	В	С	D	E
1	FIPS2 🖵	Location 🔹	Mod MAT 🔻	Count <sup>®</sup> *	State 💌
2	0	UNITED STATES	0	No	=IF(C2=0,"",RIGHT(B2,2))
3	1000	ALABAMA	0	No	
4	1001	Autauga County, AL	1	Yes	AL
5	1003	Baldwin County, AL	3	Yes	AL
53	1099	Monroe County, AL	99	Yes	AL
54	1101	Montgomery County, AL	1	Yes	AL
70	1133	Winston County, AL	33	Yes	AL
71	2000	ALASKA	0	No	
72	2013	Aleutians East Borough, AK	13	Yes	AK

After this new variable is created, PivotTables can be used to create summaries by State.

Summary Statistics by State

Structure used for PivotTable to the left

	Average	Std Dev	
State 💌	Income	Income	Count
	53521	8694	52
AK	62611	13090	29
AL	37779	8451	67
AR	35862	6242	75
AZ	42891	6118	15
CA	55558	14046	58
CO	50853	14203	64

▼ FILTERS	III COLUMNS ∑ Values
ROWS	$\Sigma$ VALUES
State 🔻	Average of INC1101213
	StdDev of INC110213
	Count of INC110213 -

The PivotTable can be sorted by any column – simply place your cursor in the column to be sorted and select Data > Sort.

## Sort inside a PivotTable

		НС	OME	INSERT	PAGE LAYOUT	r i	ORMULAS	DATA	REVIEW	VIEW	JI
Fre	om	From	Fror	From Other t Sources *	Existing Connections	Refre	Connect Connect Propert	ies ks	Sort	Filter	Clea Reap Adv
			Get	External Data			Connections		S	iort & Filter	r
			_		f. ra		Sort By Value			?	×
B4	•		*	: ^ V	Jx 544	69.82	Sort options		Sort <u>d</u> ired	tion	
4		А		В	С		© <u>S</u> mallest	to Largest	<u>T</u> op	to Bottom	
1							Largest	to Smallest	🔘 Left	to <u>R</u> ight	
2							Summary				
3	Row	Label	s 🔻	Average Inco	<u>ne</u> Count		Sort State b	y Average I	ncome in d	escending	
4				54469.	82 5	6	order				
5	AK			62959.	39 2	8					
6	AL			37778.	72 6	7					
7	AR			35861.	81 7	5		_			
8	AZ			42891.	27 1	5			OK	Cancel	
^	<b>~</b>					•					

Average Household Income sorted from largest to smallest.

	Averag	ge Std Dev	
State	🕂 Incom	e Income	Count
NJ	72387	7 16298	21
СТ	70503	8662	8
MD	69404	4 20872	24
DC	65830	) #DIV/0!	1
RI	65333	3 9853	5
MA	64871	1 12806	14
AK	62611	1 13090	29
HI	6179:	1 7752	5
MN	52327	7 9429	87
	1004		67
PA	4921:	3 9396	67
KV	38526	5 10050	120
	30320		120
WV	38210	6782	55
AL	37779	9 8451	67
AR	35862	2 6242	75
MS	34600	) 7542	82

Place your cursor in the PivotTable. A variety of charts can be created from the Charts menu. The following pareto-type graphic displays the average household income from largest to smallest.



### Questions

- 4. What is the average median household income for the state you live in? How does this value compare to others states?
- 5. Why is there a blank state label on the graphic above? Discuss.

6. Consider the following graphics. Why would standard deviation of median household income be a better measure for income disparity than an average? Discuss.



7. The pareto-type chart below show the standard deviation of incomes from highest to lowest. Identify states that appear to have low income disparity, i.e. incomes are similar across counties. What states appear to have high income disparity?



8. Notice, DC, i.e. Washington DC, on the chart above has no standard deviation. Why is this the case? Explain.

Yes 🖵

2

## Task #1

For this task, the average income levels will be separated on a new variable called BachelorPlus Levels. This variable should be setup using the existing variable EDU685213 and with the following structure.

EDU685213 Bachelor's degree or higher, percent of persons age 25+, 2009-2013

- About 1/3 of counties have EDUC685213 less than 15, label these counties as Low •
- About 1/3 of counties have EDUC685213 greater than 20, label these counties as High •
- Label the remaining counties as Medium •

	AA	AB	AC
1	EDU635213	EDU685213	BachelorPlus Levels
2	86	28.8	=IF(AB2<15,"Low",IF(AB2>20,"High","Medium"))
3	83.1	22.6	High
4	85.6	20.9	High
5	89.1	27.7	High
6	73.7	13.4	Low
7	77.5	12.1	low

Next, use this new variable to create the following PivotTable.



Yes 🖵

Average

Income 54,045

38,074

44,689

45,937

County

BachelorPlus

Levels

High

Low

Medium

Grand Total

Fix

Order

#### Finished PivotTable Yes 🖵 County **BachelorPlus** Average Levels Income Ŧ High 54,045 Medium 44,689 Low 38,074 Grand Total 45,937

### Questions

9. Does average income increase as the proportion of residents who have a Bachelors or higher increases? If so, discuss to what degree.

# Using =AGGREGATE() and SLICER

There are limitation to what PivotTables is able to compute. The list of functions available can be found under the Summarize values By tab. Notable exceptions from this list include median or more generally percentiles. The =AGGREGATE() function and the SLICER feature in Excel can be used as an alternative to PivotTables.

Value Field Settings
Source Name: INC110213
Custom Name: INC110213
Summarize Values By Show Values As
Summarize value field by
Choose the type of calculation that you want to use to summarize data from the selected field
Sum Count Average Max Min Product Count Numbers StdDev StdDev Var Var Varp
Number Format OK Cancel

To invoke the SLICER feature in Excel, convert the dataset to an Excel Table. Give the Table a name, e.g. QuickFacts.

Tab Qui	le Name: ckFacts Resize Table	Image: Summarize with PivotTable       Image: Summarize with PivotTable         Image: Summarize with PivotTable       Image: Summa					s Browser	
Ψ	Properties	Tools		Slicer	Ŧ	External Table Data		
G9 ▼ : × √ f <sub>x</sub> 10764								
	А	В		С		D	E	F
1	FIPS2 💌	Location	-	Mod M/	ATH 💌	Count 💌	Stat 💌	fips 🔹
2	0	UNITED STATES			0	No		(
З	1000	ALABAMA			0	No		100
4	1001	Autauga County, AL			1	Yes	AL	100
5	1003	Baldwin County, AL			3	Yes	AL	100
6	1005	Barbour County, AL			5	Yes	AL	100
7	1007	Bibb County, Al			7	Yes	AI	100

To specify a SLICER, select Insert > Slicer.



The following can be used to setup a SLICER for State.

In the Insert Slicers window, specify the field (or fields) from which to construct the slicer	If State is speci following windo	fied, then the w is displayed
	State	<b>*</b>
Insert Slicers	АК	<u> </u>
FIPS2	AL	
☐ County ♥ State	AR	
fips PST045214 PST045213	AZ	
PST040210 PST120214 PST1 20212	CA	
POP010210 AGE135213	со	
AGE295213 AGE775213 SEX255213	СТ	
RHI125213 RHI225213	DC	•
RHI325213 RHI425213		

In an empty column, enter a sequence of values from 0 to 1 by increments of 0.1. These will be used to compute percentiles for income.

# Cell F2: =PERCENTILE( QuickFacts[INC110213], E2 )

Enter this =PERCENTILE() function into cell F2 as shown. Copy this down for the remaining cells.

	А	В	С	D	E	F	G	H	Ι	
1					Perce	ntiles				
2	State	)		¥×	0	= PERCEN	TILE( Quick	Facts [ INC		E2)
3	AK			<b>^</b>	0.1	33394				
4				E	0.2	36802.4				
5	AL				0.3	39592				
6	AR				0.4	42015.8				
7	47				0.5	44301				
8	AZ				0.6	46773.8				
9	CA				0.7	49651.4				
10	0				0.8	53238.6				
11				$\leq$	0.9	60477.4				
12	СТ				1	122238				
13	<b>D</b> C									

ck AK to ge	t percentile	s for S Percer	tate = A ntiles	Click AL	to get it's p	ercenti Percer	iles ntiles
State	T <sub>×</sub>	0	19986	State	™~	0	19986
АК	<b>^</b>	0.1	33394	ΔΚ		0.1	33394
,	=	0.2	36802.4		=	0.2	36802.4
AL		0.3	39592	AL		0.3	39592
AR		0.4	42015.8	AR		0.4	42015.8
47		0.5	44301			0.5	44301
AZ		0.6	46773.8	AZ		0.6	46773.8
CA		0.7	49651.4	CA		0.7	49651.4
CO		0.8	53238.6	60		0.8	53238.6
		0.9	60477.4			0.9	60477.4
СТ		1	122238	СТ		1	122238

The =AGGREGATE() function in Excel should be used in Excel when certain rows should be excluded from the requested calculations. =AGGREGATE() is a more complete version of the =COUNTIF() functions used in the previous handout.

A explanation of the required arguments for the =AGGREGATE() function are briefly discussed here.

=AGGREGATE(function\_num, options, array, [k])

The function\_num is a number from the following list. 16 should be used for percentiles

=AGGREGATE(	
AGGREGATE(function	_num, options, array, [k])
AGGREGATE(function	1 - AVERAGE
	🖾 2 - COUNT
	🖾 3 - COUNTA
	🖾 4 - MAX
	🖾 5 - MIN
	6 - PRODUCT
	7 - STDEV.S
	🖂 8 - STDEV.P
	🖾 9 - SUM
	⊡ 10 - VAR.S
	🖾 11 - VAR.P
	12 - MEDIAN
	13 - MODE SNGL
	14 - LARGE
	<u></u>
-	16 - PERCENTILE.INC
	17 OLAPTIE INC
	18 - PERCENTILE.EXC
	19 - OUARTILE.EXC

The options value should be selected from the following list. 5 will ignore hidden rows in it's calcuations.

n 🐼 o	) - Ignore nested SUBTOTAL and AGGREGATE functions
6.) 1	- Ignore hidden rows, nested SUBTOTAL and AGGREGATE functions
() 2	2 - Ignore error values, nested SUBTOTAL and AGGREGATE functions
6.) 3	- Ignore hidden rows, error values, nested SUBTOTAL and AGGREGATE functions
	5 - Ignore hidden rows
7	7 - Ignore hidden rows and error values

Replace the =PERCENTILE() function used above with the following function.

# Cell F2: =AGGREGATE(16, 5,QuickFacts[INC110213],E2)

Copy this down for the remaining percentiles.



The slicer can be used to specify any state or a collection of states.



# <u>Task #2</u>

\_

For this task, apply a SLICER on the median household income, i.e. variable INC110213. Use the slicer to verify that the following 5 counties in United States have the highest and lowest median household income.

	Counties w	vith highest median incom	e
1	FIPS2 🚽	Location 🗾	
1229	24027	Howard County, MD	
1845	35028	Los Alamos County, NM	
2898	51059	Fairfax County, VA	
2922	51107	Loudoun County, VA	
2976	51610	Falls Church city, VA	

1 FIPS2 Location	
	1
430 13061 Clay County, GA	430
1039 21051 Clay County, KY	1039
1087 21147 McCreary County, KY	1087
1108 21189 Owsley County, KY	1108
2593 48047 Brooks County, TX	2593

Questions

- 10. What are the median household income values for each set of counties listed above?
- 11. There are 3143 counties, so the richest 31 counties would represent the top 1%. How many of the top 31 counties are from VA? How about MD?

# <u>Task #3</u>

There are seven counties surrounding the Twin Cities that are known locally as the "metro counties." A map of these counties is provided here.



Create necessary variables and then use PivotTables to find the average median household income for Metro = No and Metro = Yes counties in MN. How does the average income compare across these two geographic regions? Discuss.