Review of Visualizations in Federal Statistical Agency Publications

Vicki Lancaster, Ph.D. Principal Scientist
val7zv@virginia.edu

Samantha Cohen, Ph.D. Postdoctoral Research Associate
samcohen@virginia.edu

University of Virginia
Biocomplexity Institute and Initiative
Social and Decision Analytics Division

October 30, 2019

Introduction

This is the third of a series of three papers, the first two provide comments and suggestions on the visualizations appearing in NCSES publications and provide alternative ways to display the data; the third is a survey of the visualizations in other federal government statistical agency publications.

This third paper is a survey of the visualizations in other federal government statistical agency publications. The majority of federal statistical agency visualizations are a form of bar chart (simple, stacked, multi-set), pie chart, or line graph. The exceptions are visualizations in the Bureau of Transportation Statistics and Census documents. Only those visualizations or combinations of visualizations that are not currently used in NCSES documents are included in this report. The report includes examples of these visualizations using NCSES data; these visualizations are outlined in orange for emphasis.

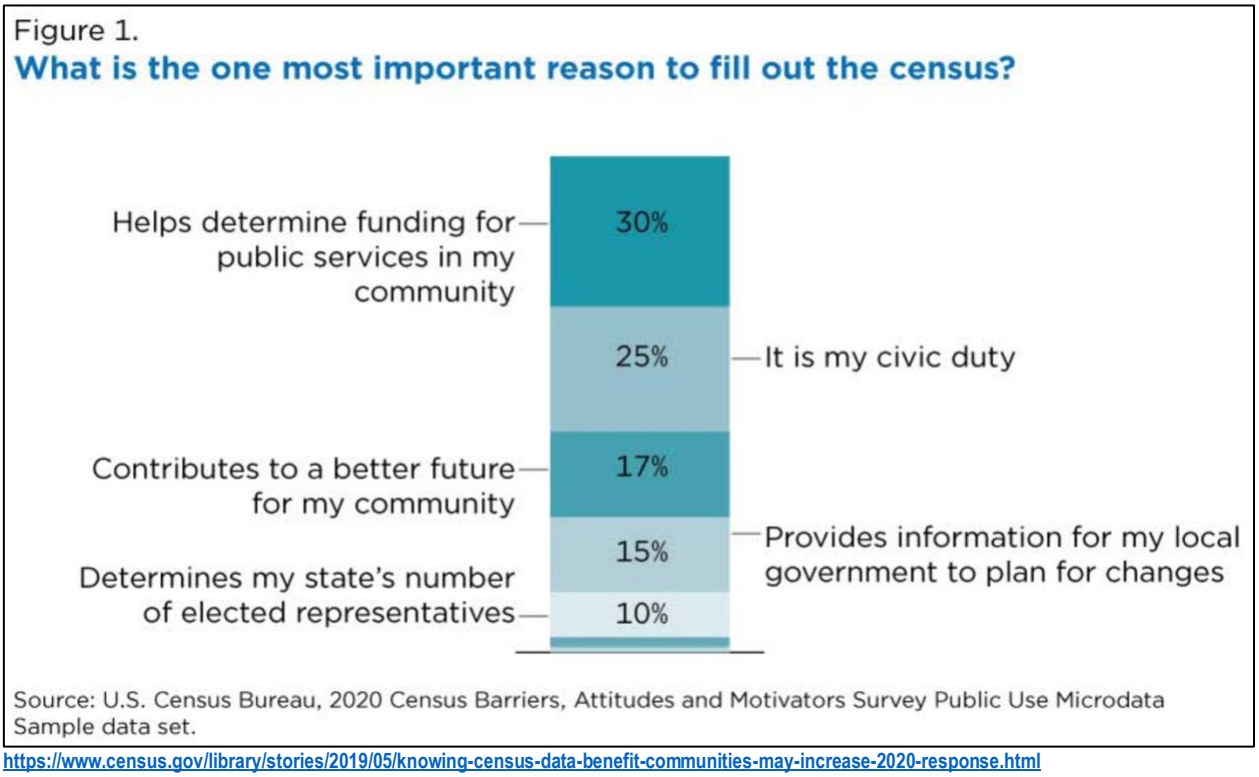
All three papers are organized by topic defined by the number of variables being displayed and the measurement level of the variables, or whether the visualization is geospatial, displays a difference between two variables, or has time on the horizontal axis. Within a topic area, the figures outlined in orange are from NCSES publications and the figures outlined in grey provide an alternative way to display the same data. The alternative visualizations presented in the paper have not been endorsed or approved by NCSES for use in NCSES publications.

Topics
1. Graphs with 1-categorical variable
2. Graphs with 2-categorical variables
3. Graphs with $\geq$ 3-categorical variables
4. Graphs with Time on the Horizontal Axis
5. Graphs that Display a Difference
6. Geospatial Graphs

This project has been funded, either wholly or in part, with federal funds from the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation (NSF) under cooperative support agreement number NCSE-1641251. The contents of this publication do not necessarily reflect the views or policies of NCSES or NSF, nor does mention of trade names, commercial products, or organizations imply endorsement of same by the U.S. Government.

1. Graphs with 1-categorial variable

Simple Stacked Bar Graph



https://www.census.gov/library/stories/2019/05/knowing-census-data-benefit-communities-may-increase-2020-response.html

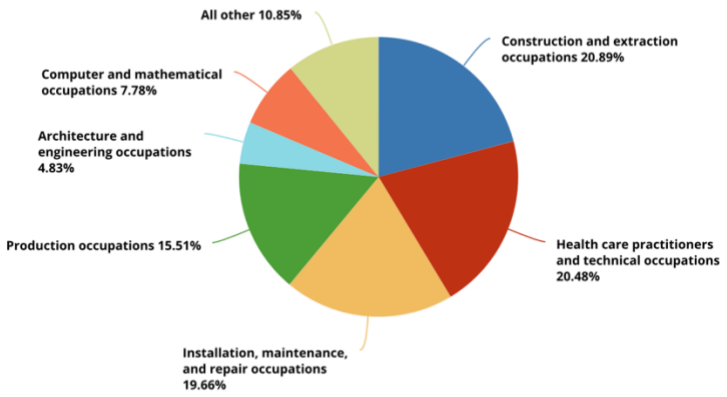
COMMENTS:

- A way to present a categorial variable with levels that require a description of more than a few words. In addition to displaying the numbers, it can also be used to empathize the information described in the body of the text. In this case Census is visualizing the results of the survey question in an effort to increase participation in the 2020 census and drive home the benefits to the community.
- A drawback of this figure is that the same colors are used to identify more than one level (for example the color for 30% and 17% are the same); a rule-of-thumb is to use a different color for each level of a categorical variable.

Example of an annotated stack-bar chart using NCSES data.

FIGURE 3-17

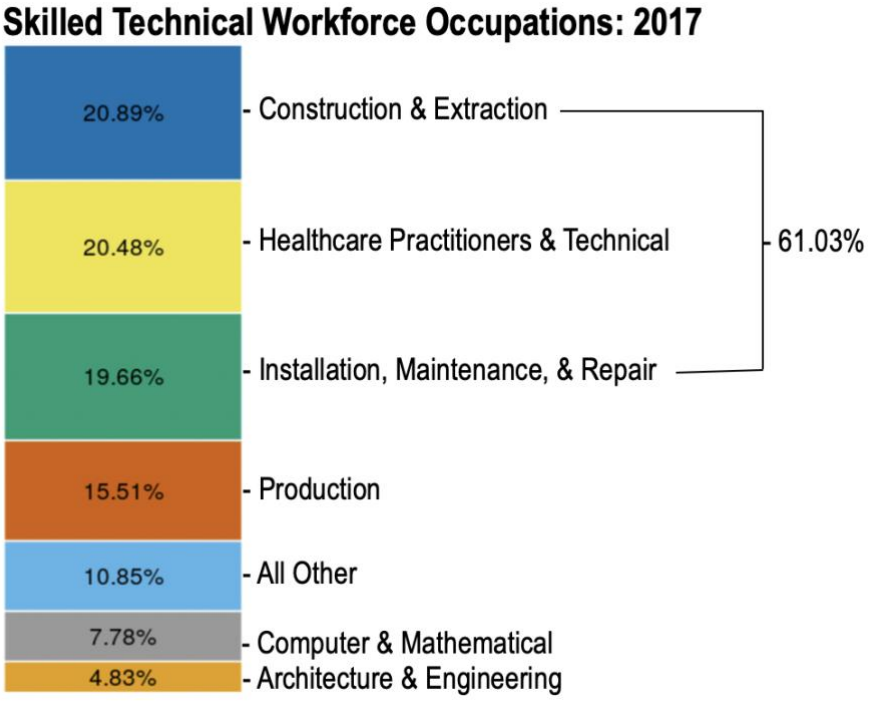
Skilled technical workers, by occupation: 2017



Note(s)
Values do not include those employed in military occupations.

Source(s)
Census Bureau, American Community Survey (ACS), 2017, Public Use Microdata Sample (PUMS).

"Employment of skilled technical workers was concentrated in four broad occupational categories. Occupations of the STW include S&E and S&E-related occupations, but also those that require significant technological skills and expertise, but do not necessarily require a bachelor's degree for entry (for a full list of occupations included here see Technical Appendix). Nearly two-thirds of the STW were employed in occupations in construction and extraction (21%); health care (20%); and installation, maintenance, and repair (20%). Another 16% were employed in production occupations (Figure 3-17)."



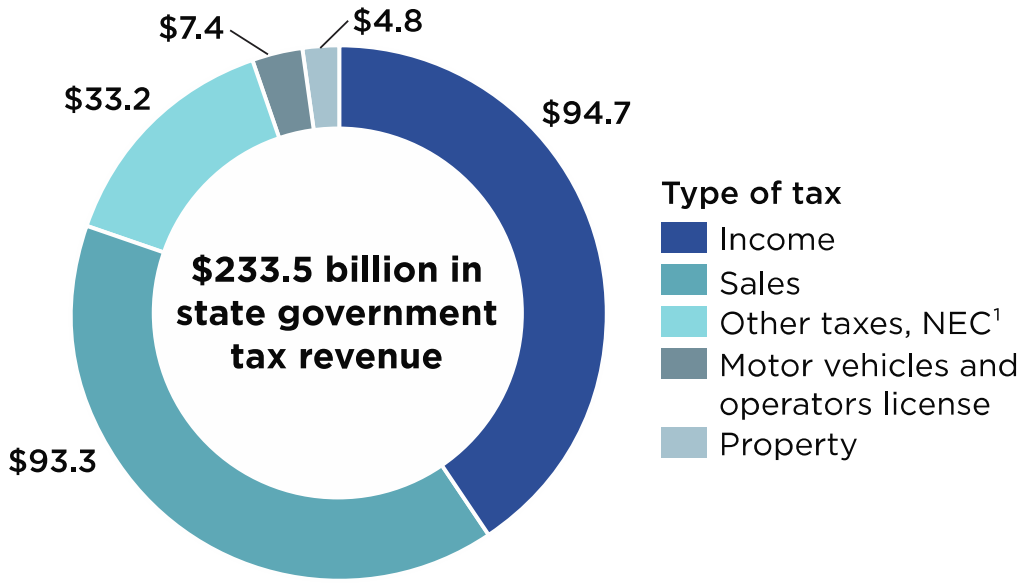
This figure emphasizes the comparisons made in the above text.

Simple Donut Chart

Figure 2.

National State Government Tax Revenue Totals: 2018 Q4

(In billions)



[1] Not Elsewhere Classified (NEC).

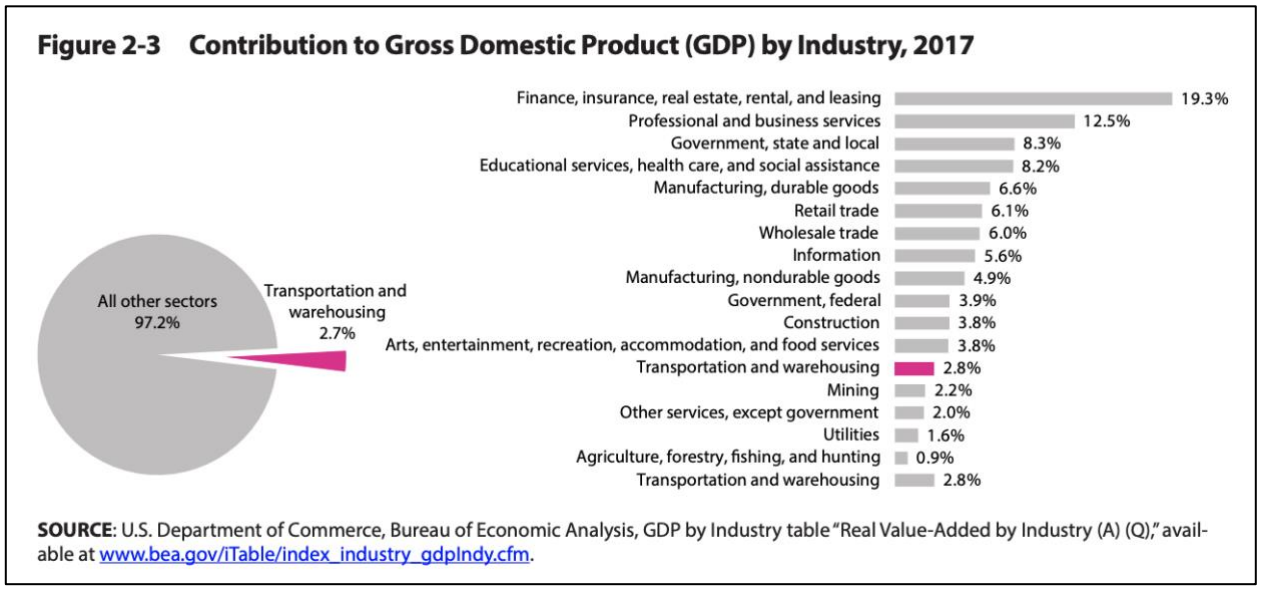
Source: U.S. Census Bureau, Quarterly Summary of State and Local Government Tax Revenue.

https://www.census.gov/content/dam/Census/library/publications/2018/econ/g18-qtax4.pdf

COMMENTS:

- This is an alternative to the stacked bar chart displayed on page 2 and functions in the same way, it emphasizes a particular variable that otherwise might get lost in the body of the text.
- Pie versus Donut Charts: the area inside the doughnut can be used to display additional information in this case the total tax revenue, otherwise pie and doughnut charts both suffer from perceptual problems, whether an arc or an angle it is hard for humans to interpret data displayed in this form.
- When the arcs are labeled with the quantity they represent it eliminates the perceptual problems and provides an alternative to simple or stacked-bar plots.
- A drawback of this visualization is that the colors used to identify the levels are so similar in some cases that it is hard to make the connection between the colors in the legend and the colors in the doughnut chart.

Multiple Visualizations: Exploding Pie Chart and Bar Chart



https://www.bts.gov/sites/bts.dot.gov/files/u796/TET%202018.pdf

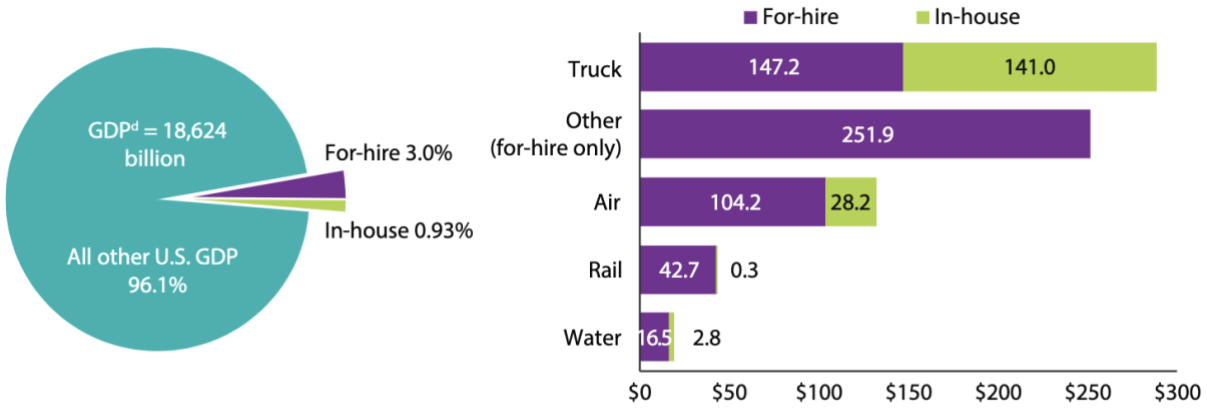
COMMENTS:

- A way to emphasize a single level of a category and compare it to the whole.
- In addition to the exploding pie chart, the contrasting colors emphasizes the variable of interest.
- Although the opinions regarding pie charts tend to the negative (do not use), in this case using it to emphasize one level of a categorical variable – Transportation and Warehousing – and comparing it to the remaining levels works. The exploding pie chart makes a greater visual impact than the bar chart.
- There is no information in the document text why Transportation and Warehousing is plotted twice in the bar chart and why the value in the pie chart and bar chart do not match.

2. Graphs with 2-categorical variables

Multiple Visualizations: Pie Chart and Stacked Bar Chart

Figure 1-1 Contribution of For-Hire and Business-Related In-House Transportation Activity to U.S. Gross Domestic Product (GDP), 2016 (current dollars)



NOTES: (a) In-house transportation is business-related transportation. Business-related transportation includes privately owned and operated vehicles of all body types, used primarily on public rights of way, and the supportive services to store, maintain, and operate those vehicles. (b) For-hire transportation consists of the services provided by transportation firms to industries and the public on a fee-basis. (c) Other for-hire transportation includes: pipeline, transit and ground passenger transportation, including State and local government passenger transit; sightseeing transportation and transportation support; courier and messenger services; and warehousing and storage). (d) The TSAs also show the contribution of transportation carried out by households through the use of their private motor vehicles (known as household production of transportation services (HPTS). The contribution of HPTS is not shown in the figure. For more information, see: https://www.bts.gov/transportation_satellite_accounts

https://rosap.ntl.bts.gov/view/dot/36436

COMMENTS:

- A way to make comparisons to the whole for two levels of a categorical variable (in-house and for-hire) and provide more detailed information on both of them using a stacked bar chart.

Example of a Multiple Visualizations: Exploding Pie Chart and Bar Chart using NCSES data.

Science and Engineering Indicators 2020
NSB-2019-7 | 2019

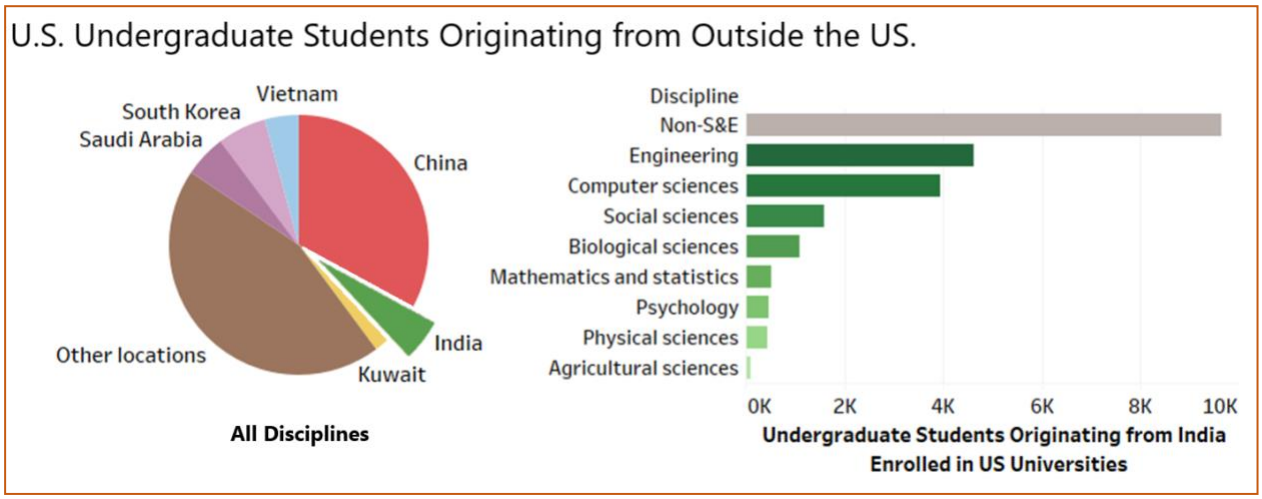
TABLE S2-13

International undergraduate student enrollment in U.S. universities, by field and selected places of origin: 2016–18

(Number)

Place of origin	All fields	S&E									Non-S&E
		Total	Engineering	Agricultural sciences	Biological sciences	Computer sciences	Mathematics and statistics	Physical sciences	Psychology	Social sciences	
2018											
All locations	435,260	179,440	56,960	3,150	17,420	35,400	16,660	8,370	9,990	31,490	255,830
China	143,310	63,480	12,890	1,390	4,120	11,710	12,090	3,520	3,300	14,470	79,830
India	22,410	12,780	4,620	100	1,100	3,940	510	440	480	1,600	9,630
Saudi Arabia	23,150	10,780	6,970	40	580	1,880	110	450	160	600	12,360
South Korea	26,190	9,420	2,300	160	1,060	2,010	750	510	710	1,920	16,770
Vietnam	18,390	6,430	1,730	70	1,090	1,960	330	450	250	550	11,960
Kuwait	7,730	6,050	5,430	0	180	50	40	80	30	230	1,680
Nepal	8,150	4,430	970	40	480	2,390	160	160	70	160	3,730
Canada	13,110	3,830	930	100	870	400	160	160	430	790	9,270
Nigeria	6,450	3,130	1,200	20	660	700	50	90	100	310	3,320
Malaysia	4,890	3,000	1,350	70	270	350	260	200	220	300	1,890
Brazil	9,210	2,410	850	40	290	400	60	80	220	480	6,800
Taiwan	7,220	2,240	530	70	290	480	160	100	210	400	4,970
Japan	9,170	2,230	270	60	210	370	100	120	260	860	6,940
Indonesia	4,860	1,980	740	80	160	370	150	50	110	320	2,870
Mexico	5,380	1,800	790	40	190	230	50	90	120	300	3,570
Pakistan	3,380	1,750	640	10	120	500	80	50	50	300	1,630

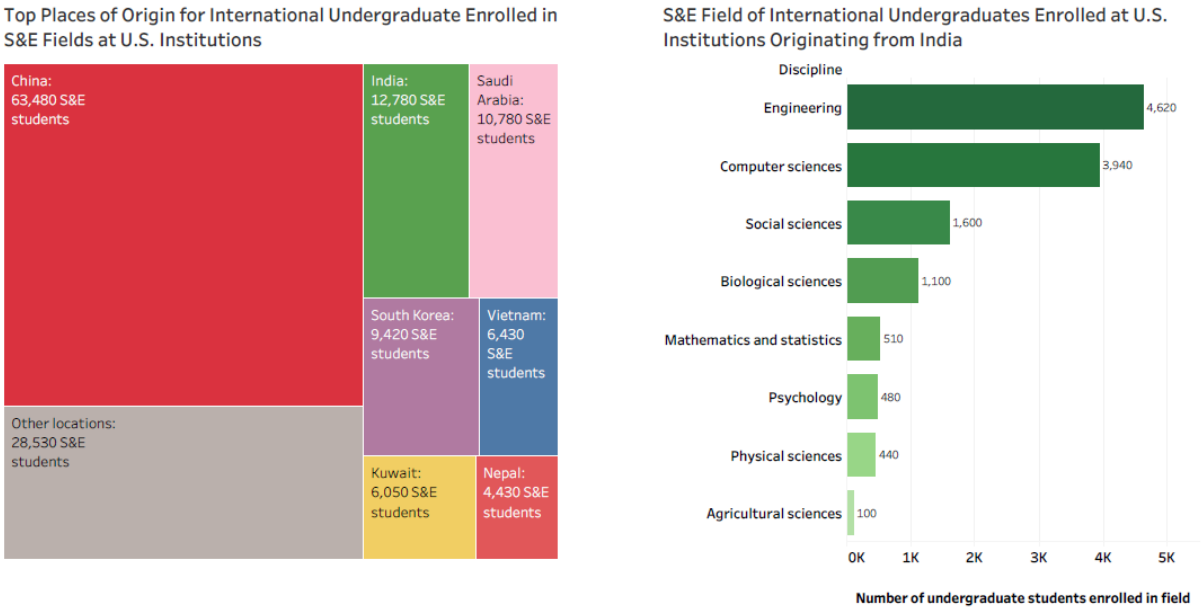
Only a portion of Table S2-12 is displayed, click on Data Tables and select Table S2-12 to view the entire table.



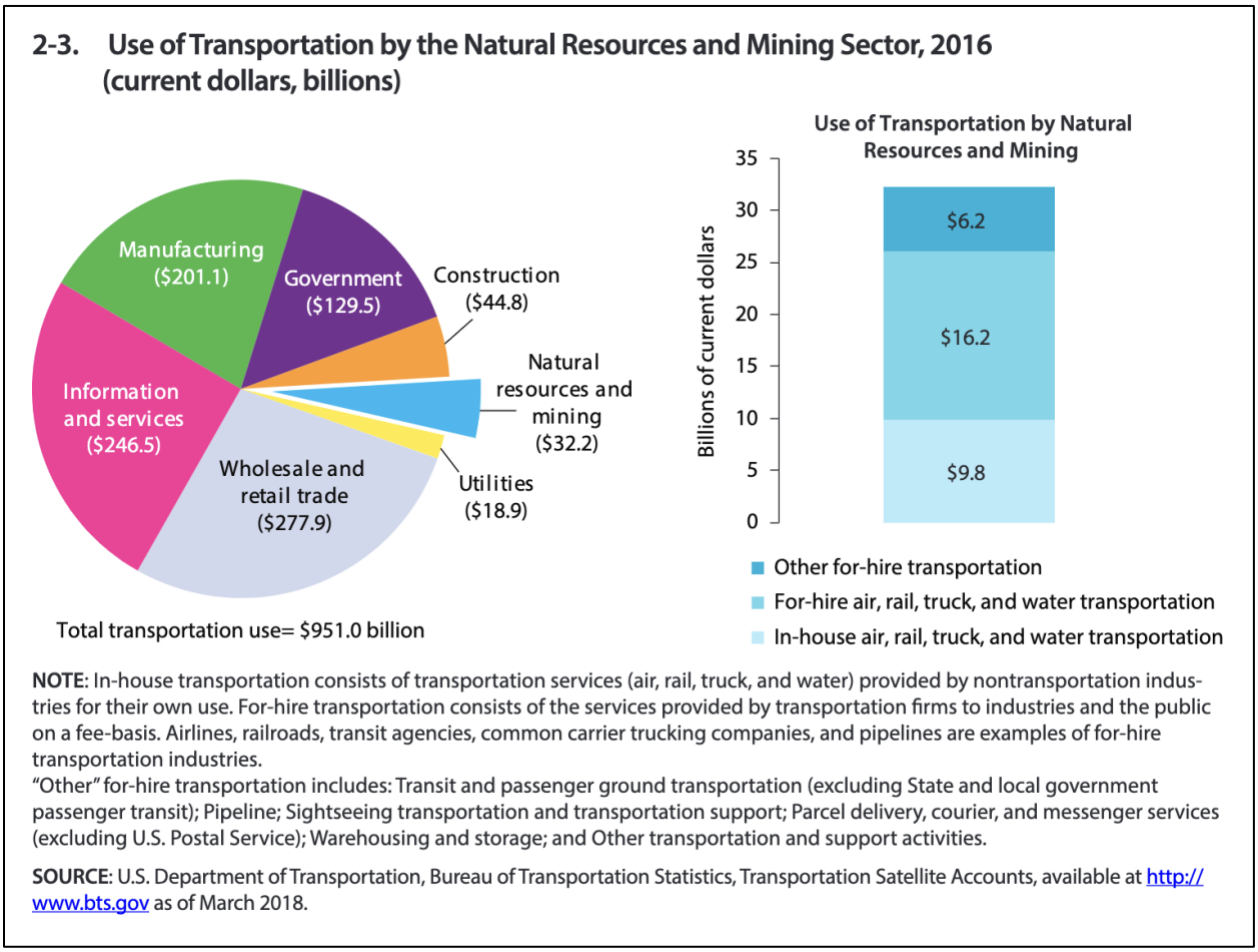
The above figure displays the top seven countries for international undergraduate students enrolled in U.S. universities in 2018 (the remaining countries of origin are collapsed into "Other locations"). The exploding pie chart shows the overall number of students from these countries for all fields of study and zooms in on India. The bar chart on the right shows the distribution of field of study for Indian students categorized by non-S&E and S&E. By pulling out this subset, we can see that amongst undergraduate students from India, Engineering and Computer Sciences account for over 50% of undergraduates.

These data are re-plotted on the next page using a mosiac plot in place of the exploding pie chart.

Using a mosaic plot instead of an exploding pie chart provides the opportunity to display additional information. The mosaic plot displays the top seven countries for international undergraduate students enrolled in S&E fields. The mosaic plot displays the name of the country and number of students enrolled in an S&E field of study in the U.S. The plot on the right displays the same data as the bar chart on the previous page but labels each bar with the number of students.



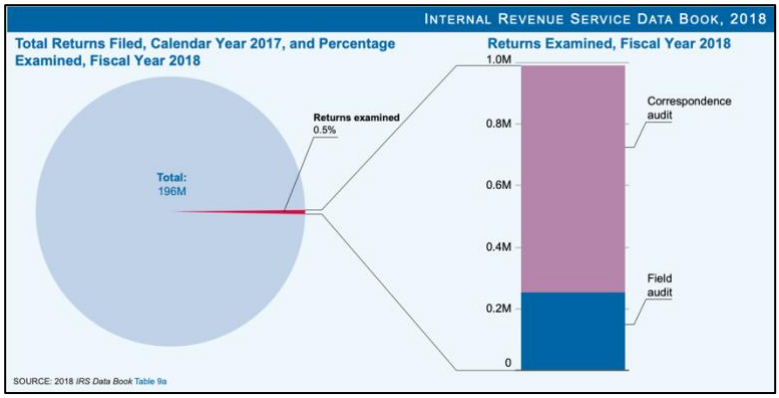
Multiple Visualizations: Pie Chart and Stacked Bar Chart



https://rosap.ntl.bts.gov/view/dot/36436#

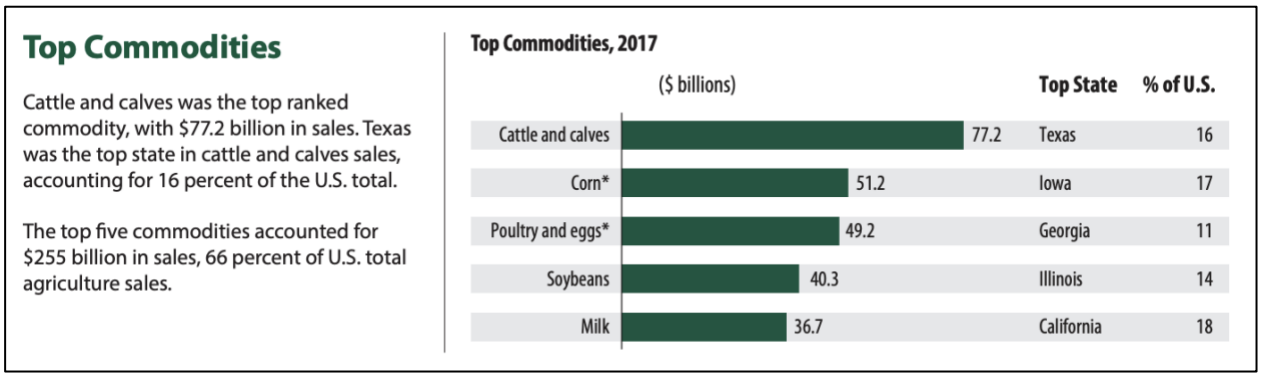
COMMENTS:

- A way to emphasize one wedge (category level) of a pie chart and introduce more detailed information on this level by including a second categorical variable using a stacked bar chart.
- An additional example from the IRS is displayed on the right which empahsizes how few tax returns are examined.



https://www.irs.gov/pub/irs-soi/18databk.pdf

Bar Chart

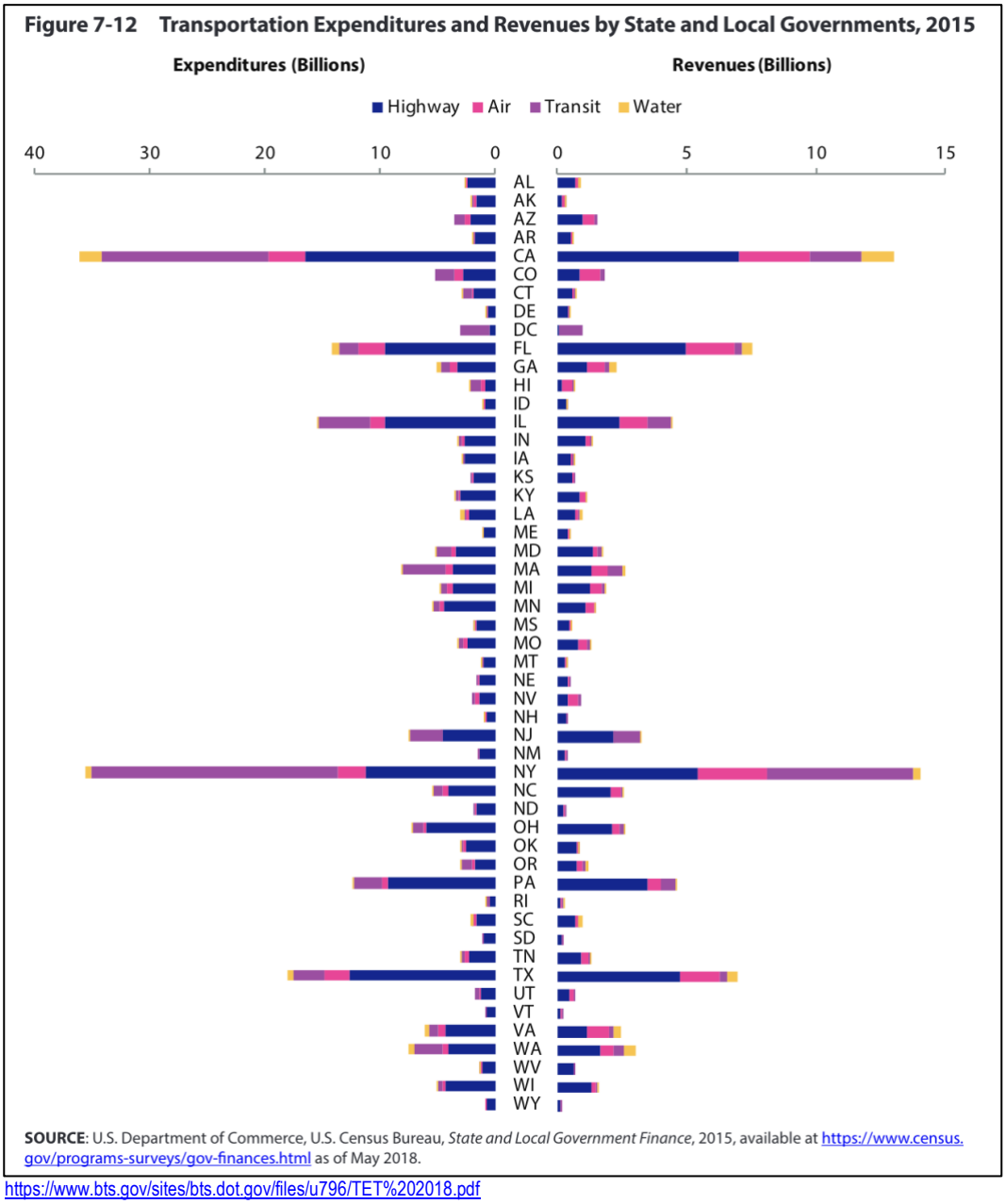


https://www.nass.usda.gov/Publications/Highlights/2019/2017Census_Farm_Economics.pdf

COMMENTS:

- A way to introduce an additional categorical variable to a bar chart. In this case, only one level of the additional variable is displayed and it is a function of the level of the bar. The top state for each commodity is identifed.

Stacked Bar Chart

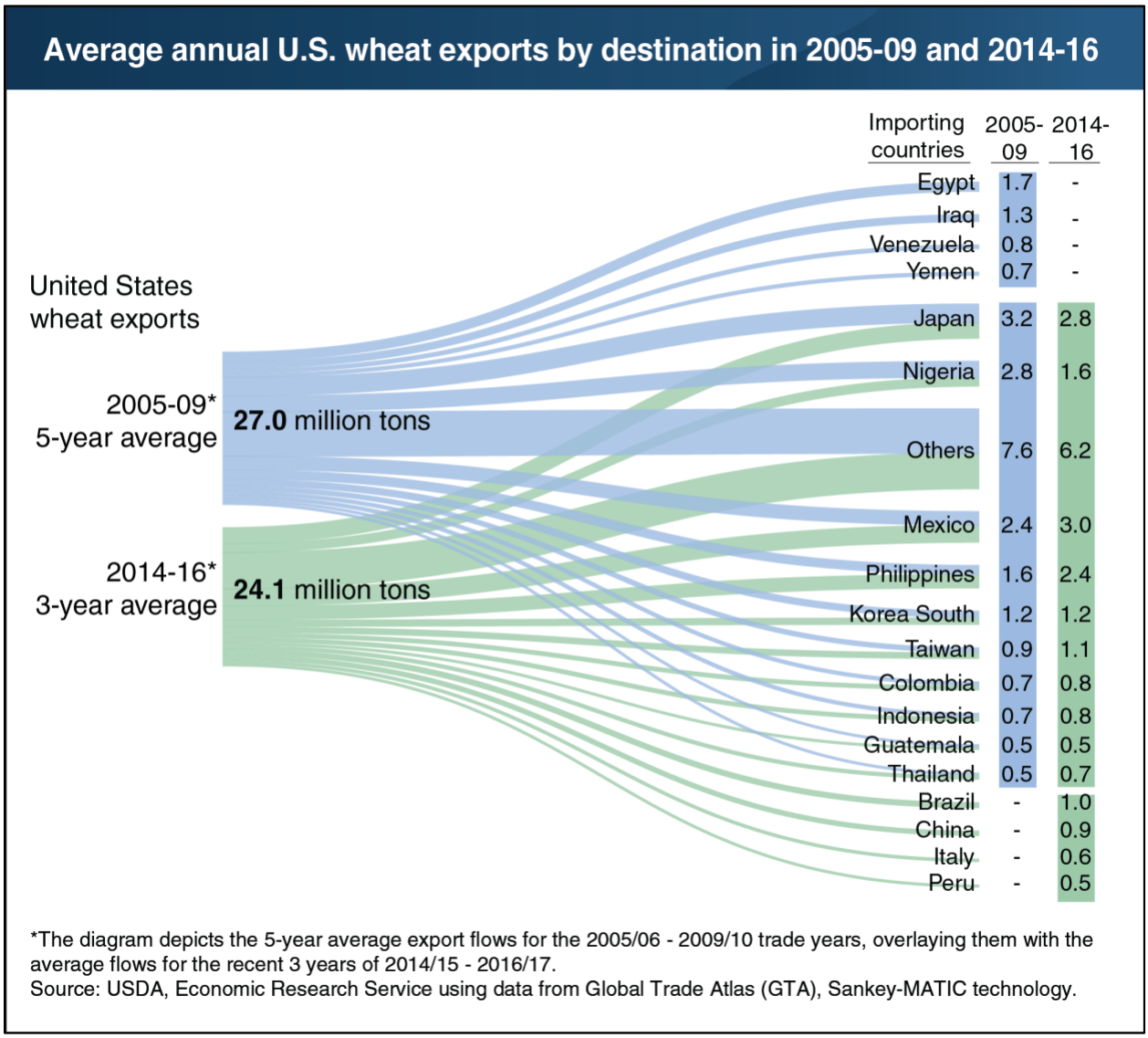


https://www.bts.gov/sites/bts.dot.gov/files/u796/TET%202018.pdf

COMMENTS:

- Could be used as a way to display information for each state for two quantitiative variables using a single categorical variable.
- Ways to make the display easier to interpret would be to include vertical grid lines – one suggestion is to use wide light gray interval grid lines for example, [0-5) white, [5-10) gray, etc.
- This display might work better if the range for the two quantitative variables was the same; here the range of expenditures (0 - ~40) is over 2x the range for revenue (0 - ~15).

Sankey Diagram

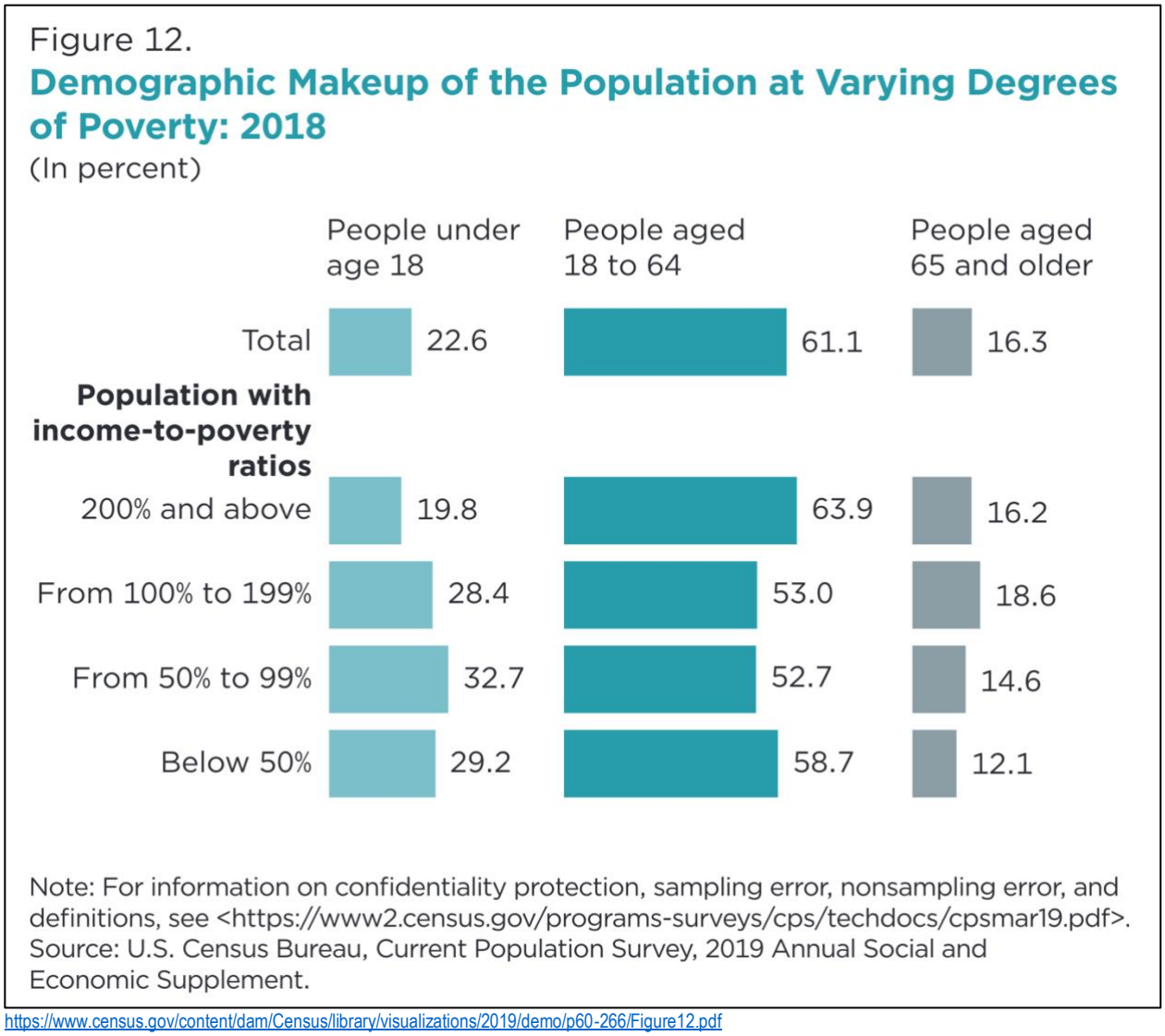


https://www.ers.usda.gov/data-products/chart-gallery/gallery/chart-detail/?chartId=90312

COMMENTS:

- A way to display a statistic (mean) over time periods for a single categoric variable.

Stacked Bar Chart with Reference



https://www.census.gov/content/dam/Census/library/visualizations/2019/demo/p60-266/Figure12.pdf

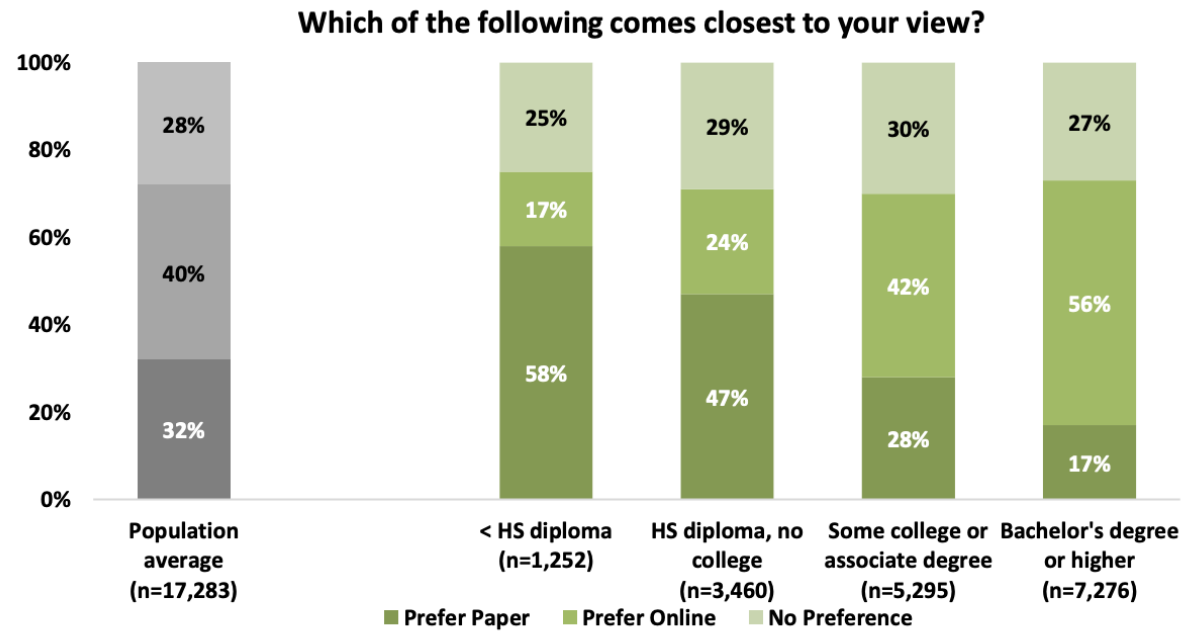
COMMENTS:

- This is an adaptation of the stacked-bar chart; the bars have been turned sideways and the individual levels separated.
- A reference set of bars is displayed on the top row.

Stacked Bar Chart with Reference

Figure 11

Less educated respondents preferred to fill out the paper census form.



Source: 2020 CBAMS Public Use Microdata Sample
Notes: (1) Unweighted sample sizes are in parentheses.
(2) Appendix A reports the standard errors for all point estimates.
(3) All estimates reported in the bar were rounded to whole numbers so that the sum of estimates equals 100%.

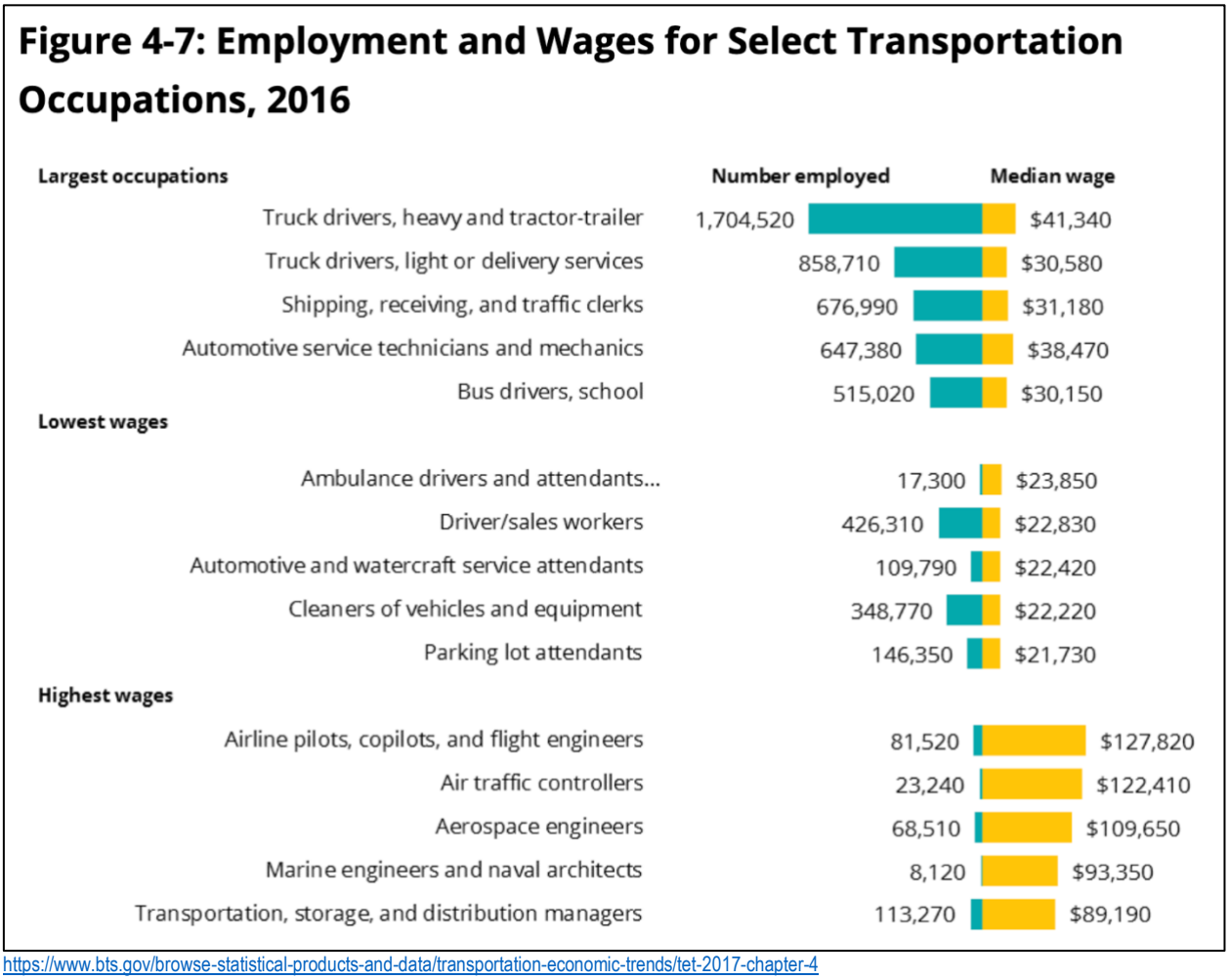
https://www2.census.gov/programs-surveys/decennial/2020/program-management/final-analysis-reports/2020-report-cbams-study-survey.pdf

COMMENTS:

- Another example of stacked bar charts which includes reference values on the left. In addition to labeling the reference bar, it is also emphasized using a different color shades.
- The addition of horizontal grid lines would make for easier comparisons between the bars.

3. Graphs with ≥ 3-categorical variables

Divergent Stacked Bar Chart

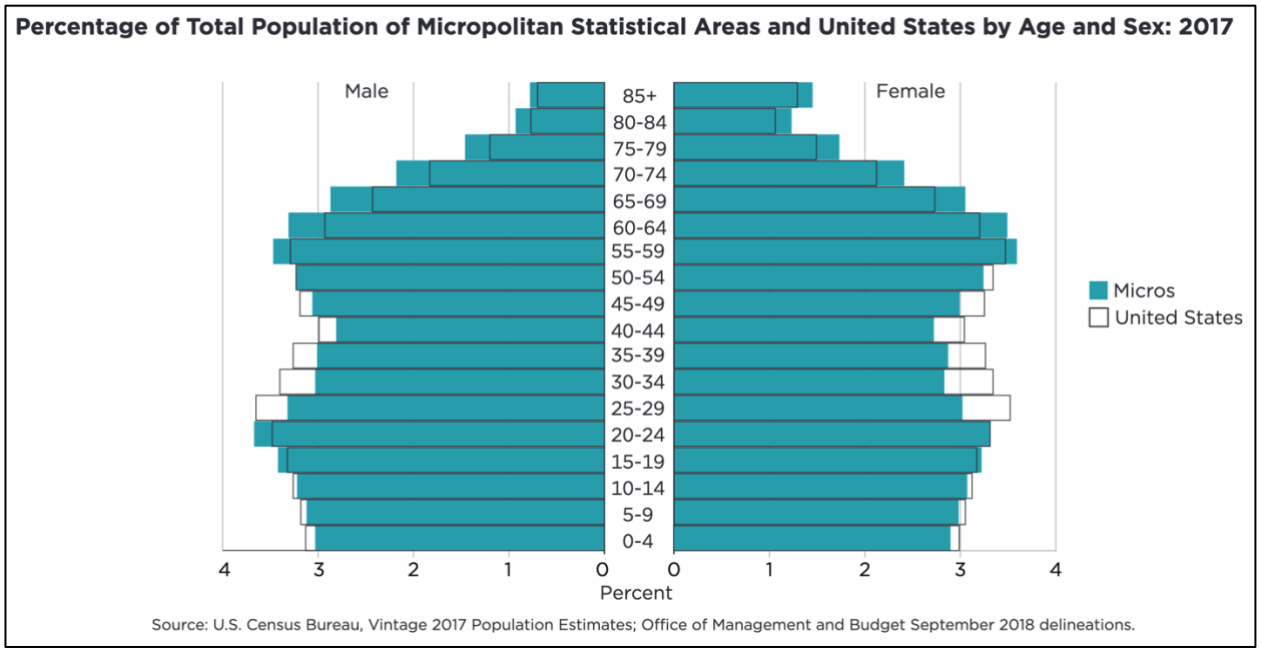


https://www.bts.gov/browse-statistical-products-and-data/transportation-economic-trends/tet-2017-chapter-4

COMMENTS:

- A way to display two quantitative variables, in this case median wage and number of employed, for occupations categorized into three discrete groups, largest occupations, occupations with the lowest wages, and occupations with the highest wages.
- Although the horizontal axis is not displayed, it is obvious that the range for median wage and number employed are different. This makes comparison between the two quantitaive variables difficult.
- A vertical line at zero would be a helpful reference between the two different scales.

Population Pyramid



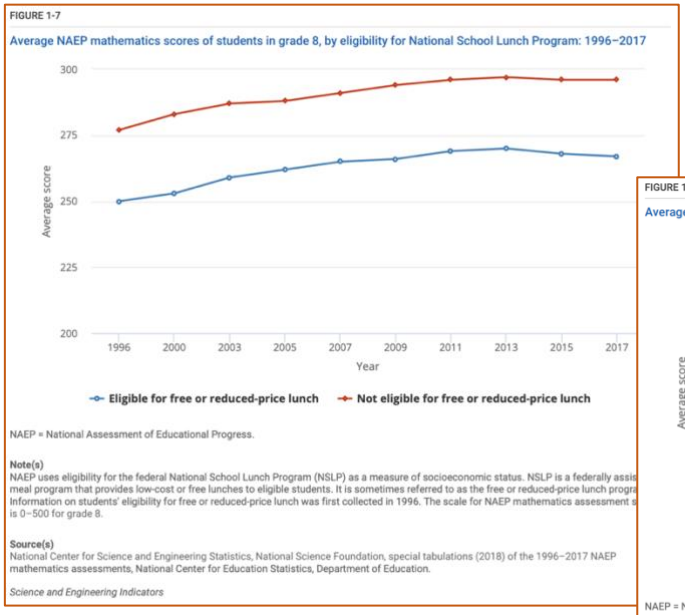
https://www.census.gov/content/dam/Census/library/visualizations/2019/demo/micropolitan-america.pdf

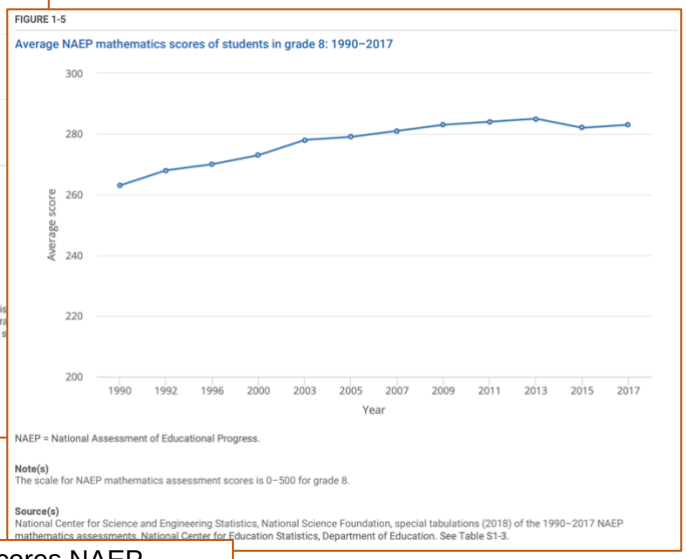
COMMENTS:

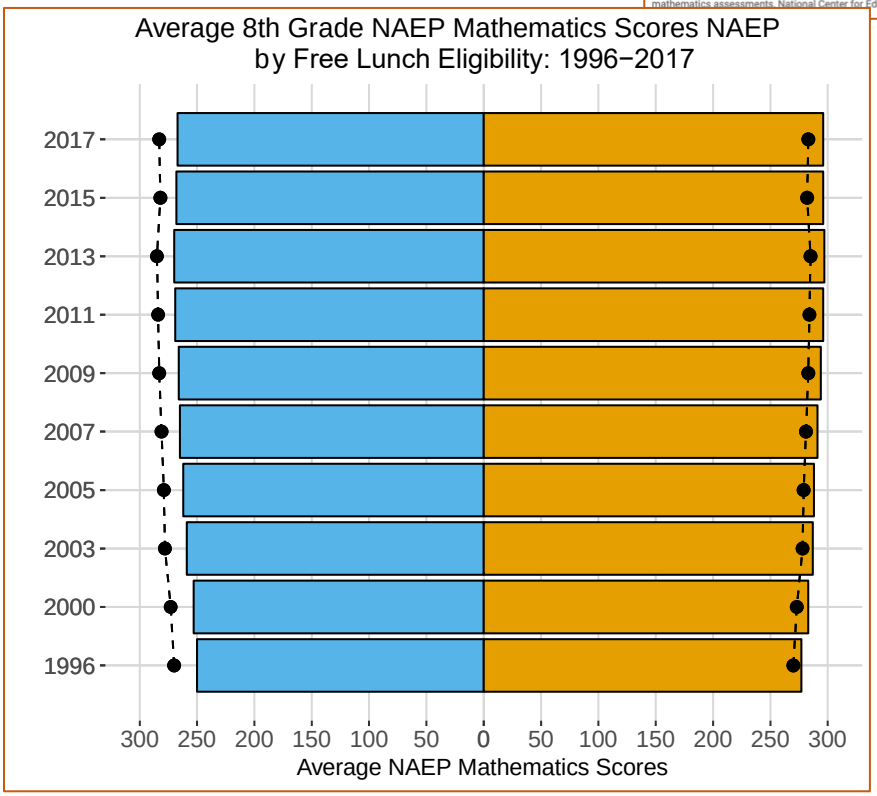
- A way to display a categorical variable with two levels, in this case gender, that have the same levels of a second variable (age categories).
- In this case the two bars are placed on top of one another by using a solid color for micro areas and a black outline for the entire U.S.

Example of a population pyramid using NCSES data.

Elementary and Secondary Mathematics and Science Education
SPECIAL REPORTS | NSB-2019-6 | SEPTEMBER 4, 2019





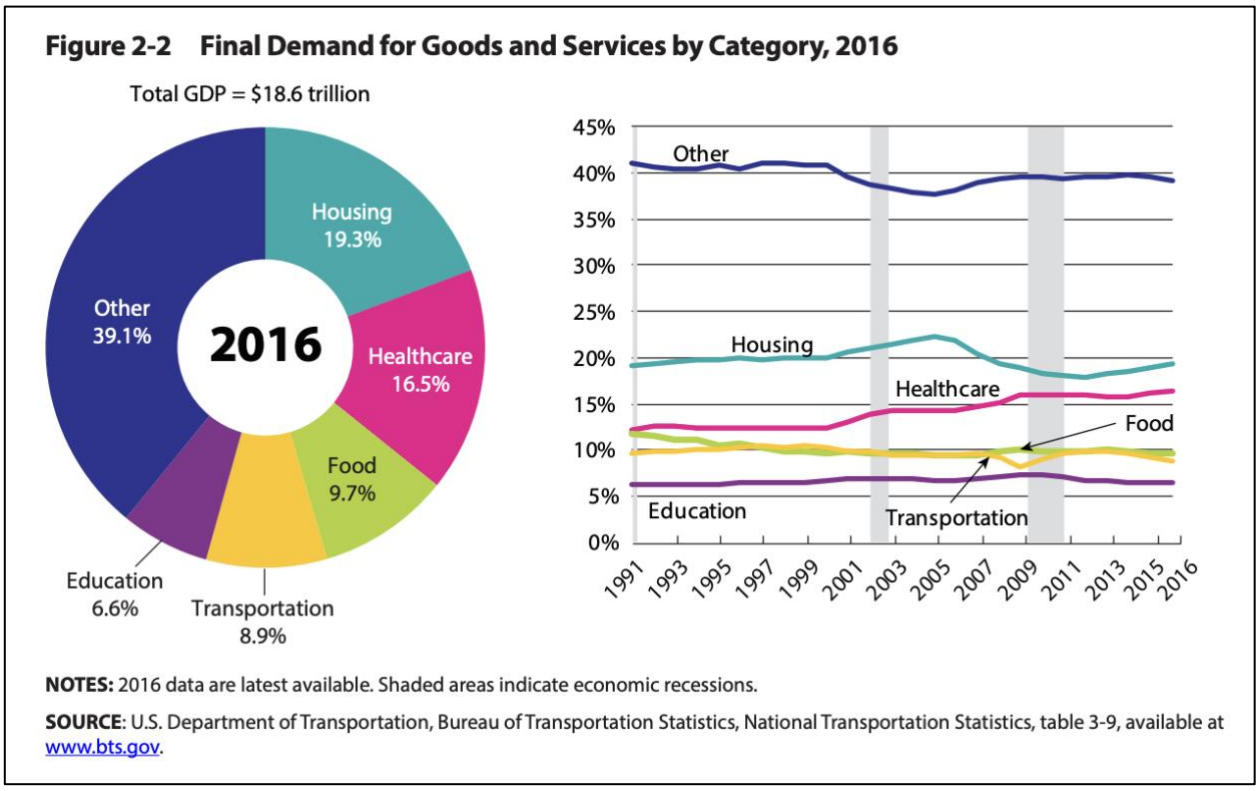


The dashed black lines are the average NAEP mathematic scores for the 8th grade.

The data from the two figures are replotted using a single figure.

4. Longitudinal Graphs

Multiple Visualizations: Simple Donut Chart and Line Graph

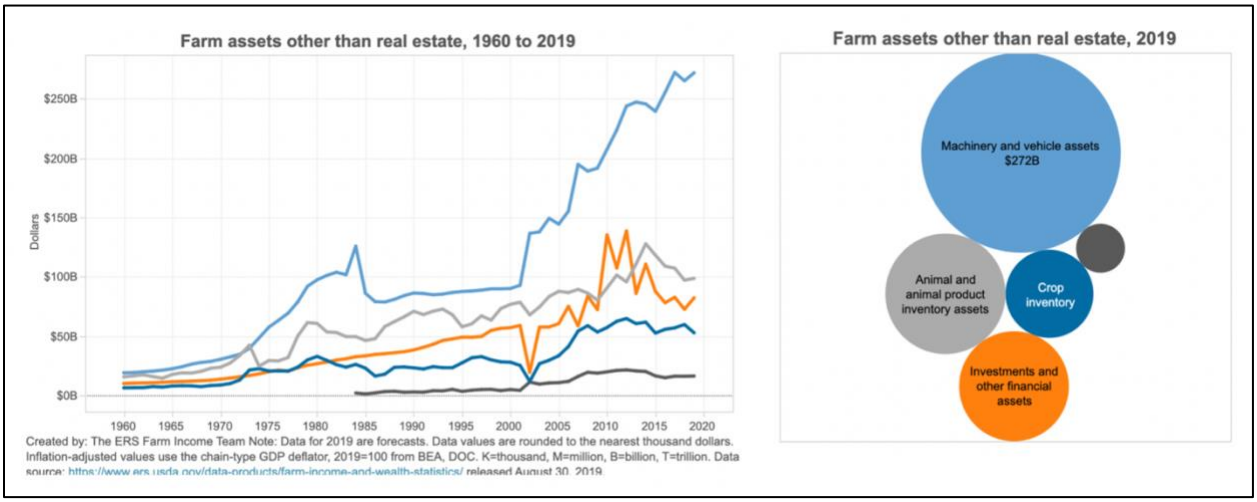


https://www.bts.gov/content/final-demand-goods-and-services-major-social-function-2016

COMMENTS:

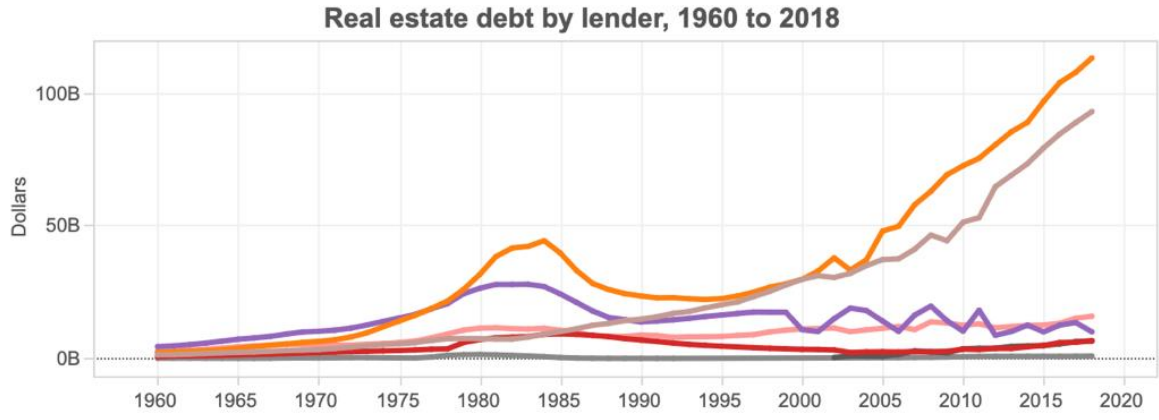
- Multiply plots within in figure – one plot shows trends over time using a line graph and data from the most recent time period is displayed separately with more detail using a donut chart.
- Outside data is displayed in the line chart to provide context (shock events, policy changes) which can also help with interpreting trends.
- Other examples of using multiple plots one of which is a line graph are displayed on the next three pages:
  - a line graph with a proportional area chart (page 19) and a line graph with a mosaic plot (page 20); and
  - in the visualization on page 21, instead of a donut chart a bar chart is used where each bar is displayed using a line chart showing trends over time.

Line Graph and Proportional Area Chart



https://public.tableau.com/views/BalanceSheetStoryPoint2/BalanceSheet?amp&:display_count=no&:embed=y&:toolbar=no&:showVizHome=no

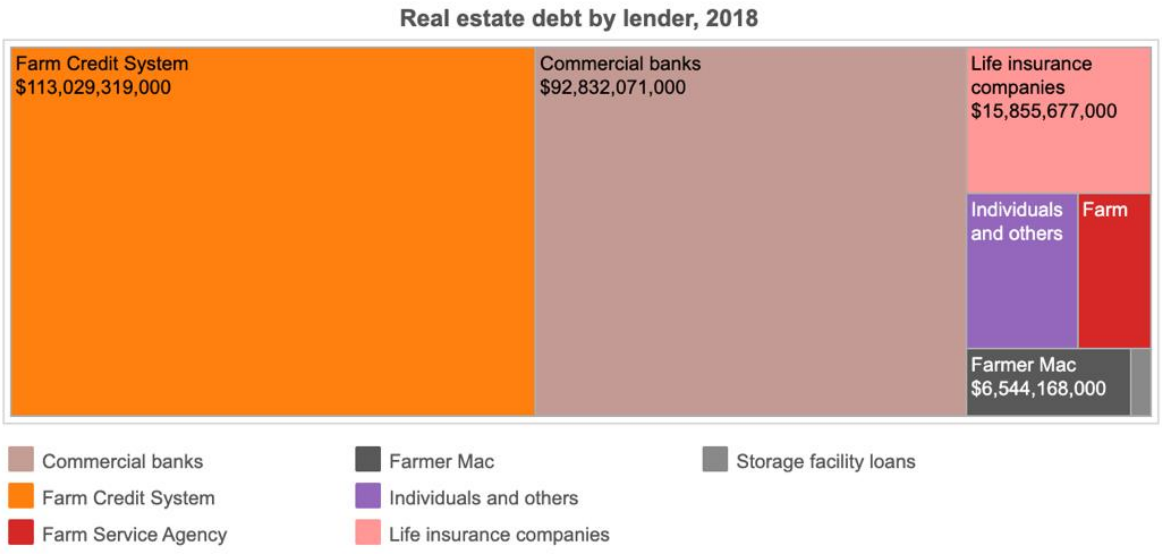
Line Graph and Mosaic Chart



Note: Data values are rounded to the nearest thousand dollars. Inflation-adjusted values use the chain-type GDP deflator, 2019=100 from BEA. DOC. K=thousand. M=million. B=billion. T=trillion.

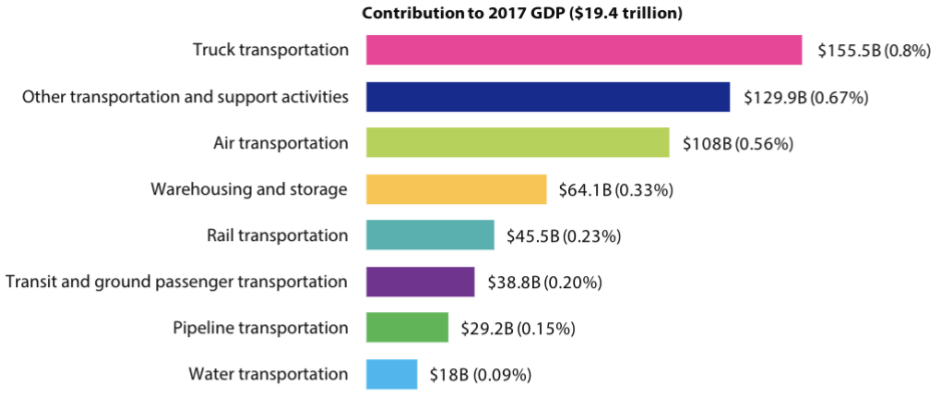
Year

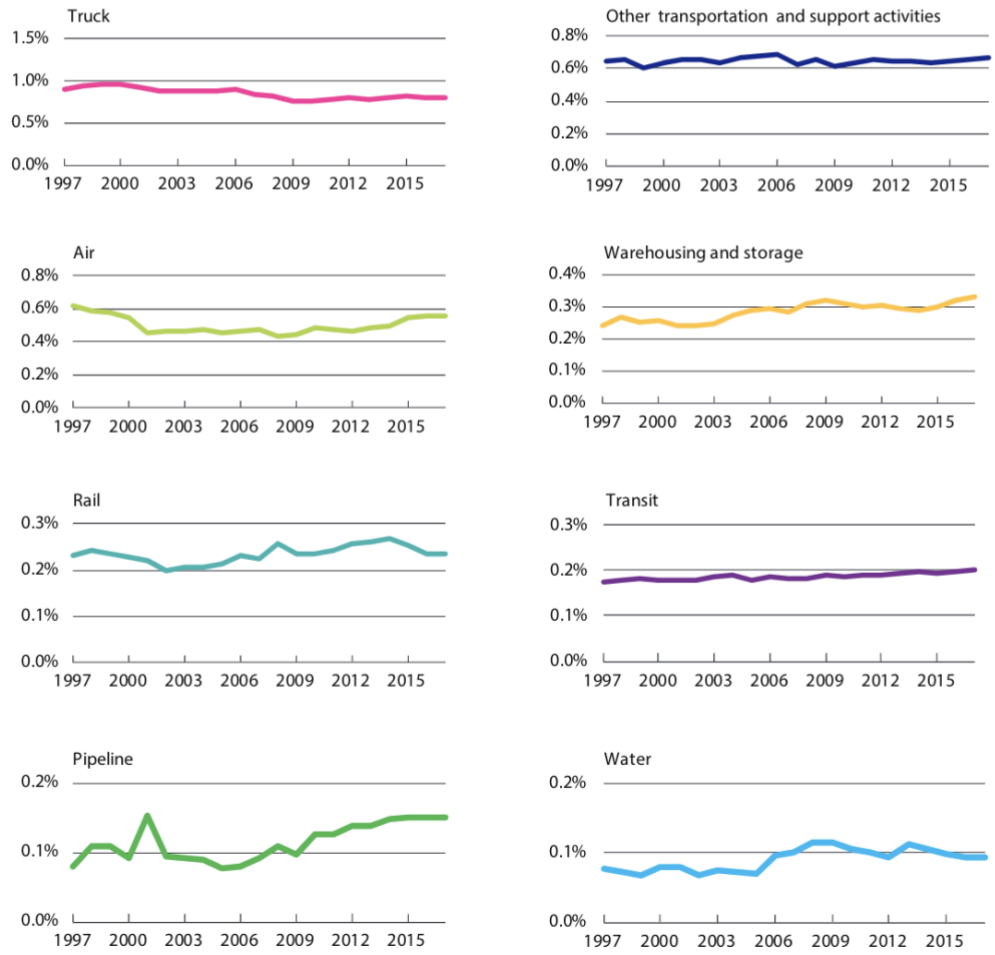
2018



Note: Beginning with 2012 estimates, farm sector debt held by savings associations is reported with the commercial bank lender group instead of the individuals and others grouping. Data source: https://www.ers.usda.gov/data-products/farm-income-and-wealth-statistics/ released August 30, 2019.

Figure 2-4 For-Hire Transportation and Warehousing's Contribution to GDP by Industry (percent)





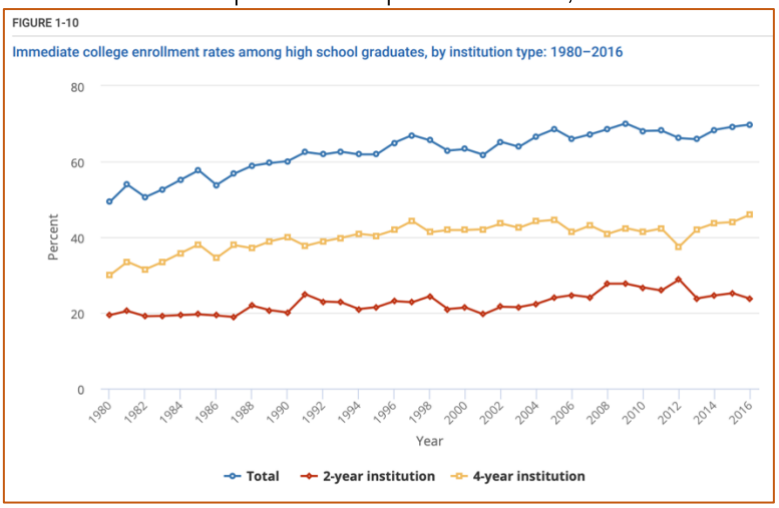
NOTES: Data are from the value-added by industry table of the BEA Industry Economic Accounts. Data for Transportation and Warehousing is Line 40, and for individual modes are in Lines 41 through 48. Current-dollar data appear in *National Transportation Statistics*, table 3-1.

SOURCE: U.S. Department of Commerce, Bureau of Economic Analysis, GDP by Industry, Value-Added by Industry Table (April 19, 2018 release), available at www.bea.gov as of June 2018.

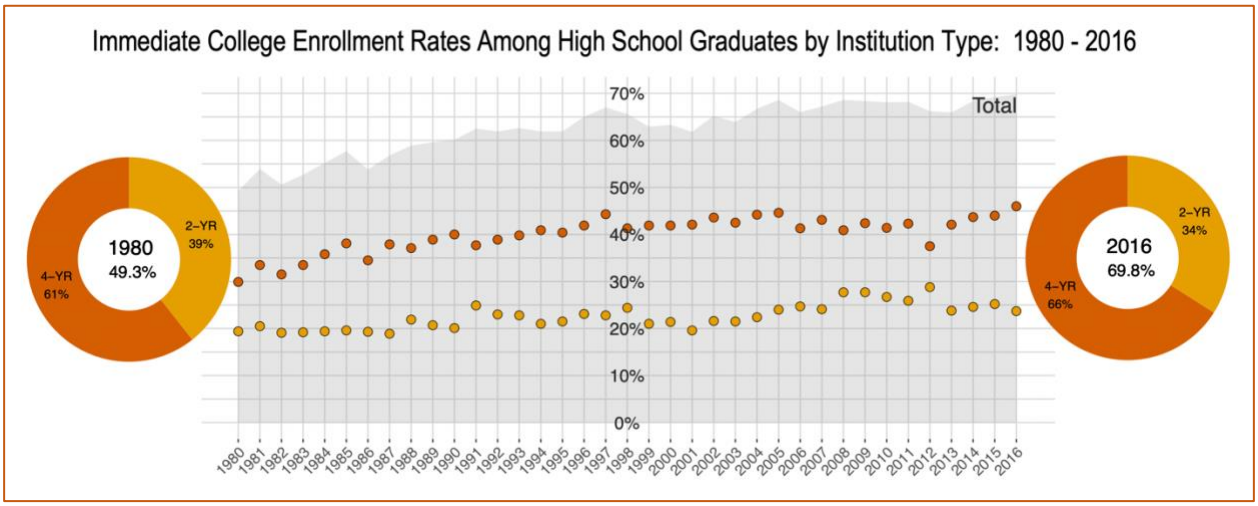
https://www.bts.gov/sites/bts.dot.gov/files/u796/TET%202018.pdf

Example of a line graph and donut chart using NCSES data.

Elementary and Secondary Mathematics and Science Education
SPECIAL REPORTS | NSB-2019-6 | SEPTEMBER 4, 2019



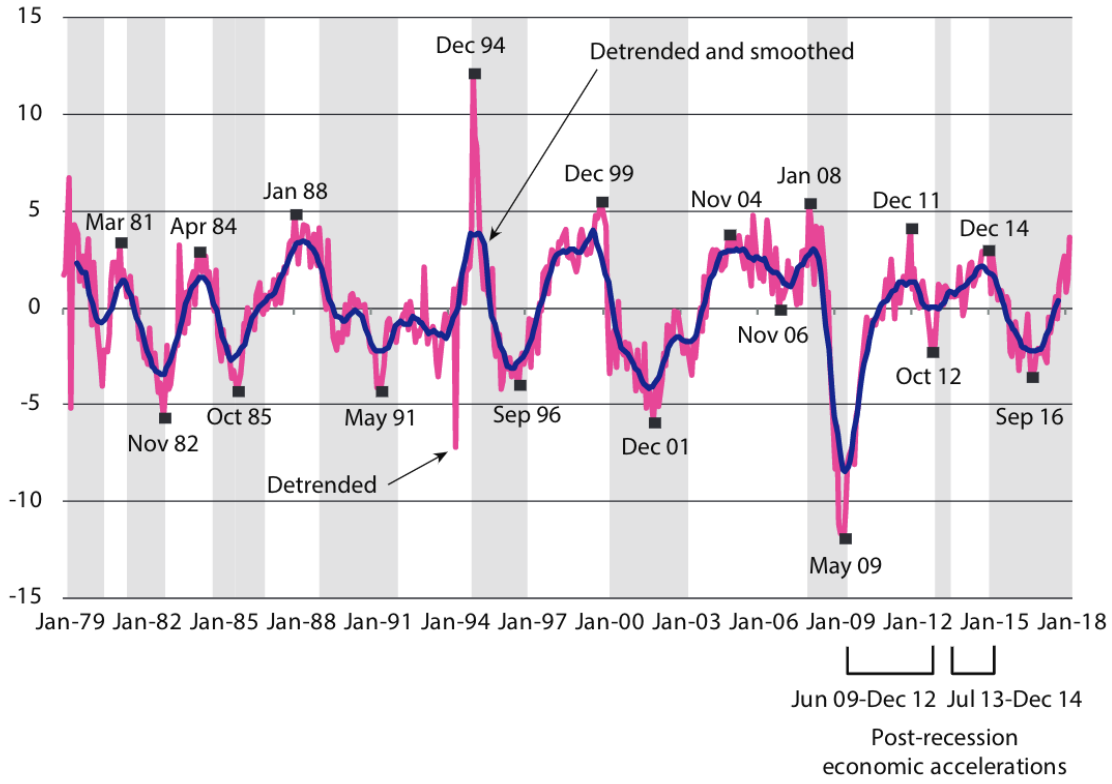
"According to data from the Current Population Survey, immediate college enrollment rates have increased over time (NCES 2019). Between 1980 and 2016, the percentage of high school graduates making an immediate transition to college increased from 49% to 70% (Figure 1-10). In addition, immediate enrollment rates rose faster between 1980 and 2016 for 4-year institutions (from 30% to 46%) than 2-year institutions (from 19% to 24%)."



This figure emphasizes comparisons made in the above text using donut plots for the first and last year of data.

Line Graph

Figure 1-3 Freight Transportation Services Index and the Economic Growth Cycle, January 1979 to March 2018



NOTES: Shaded areas indicate decelerations in the economy, and areas between are accelerations in the economy (growth cycles). Endpoint for deceleration begun in December 2014 has not been determined. Detrending and smoothing refer to statistical procedures that make it easier to observe changes in upturns and downturns of the data. Detrending removes the long-term growth trend and smoothing removes month-to-month volatility.

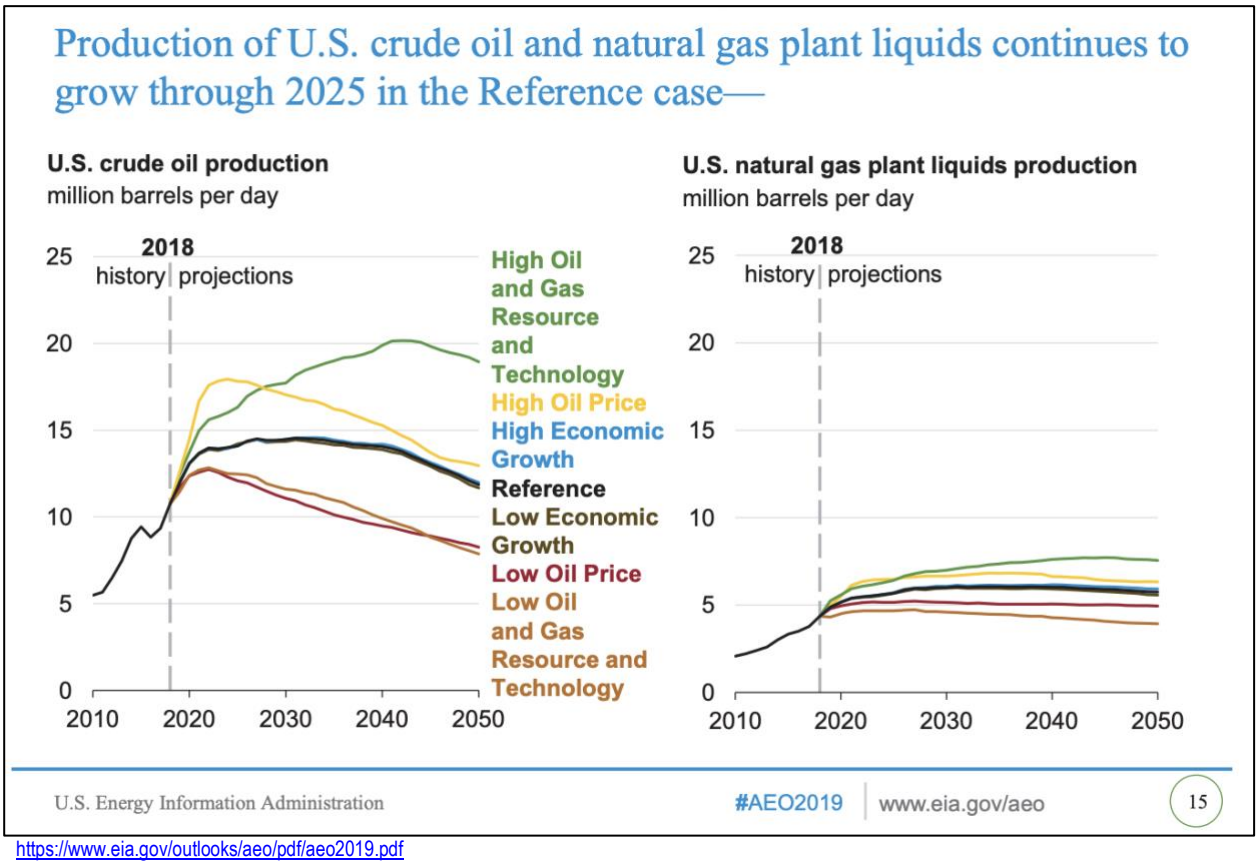
SOURCE: U.S. Department of Transportation, Bureau of Transportation Statistics, Transportation Services Index, available at www.transtats.bts.gov/OSEA/TSI as of May 2018.

https://www.bts.gov/sites/bts.dot.gov/files/u796/TET%202018.pdf

COMMENTS:

- An example of adding text to a figure to provide information; in this case low/high values within the accelerations and decelerations in the economy.
- A smoothed line is included to aid in interpretation.
- Text within the figure is used instead of a legend.

Line Graph

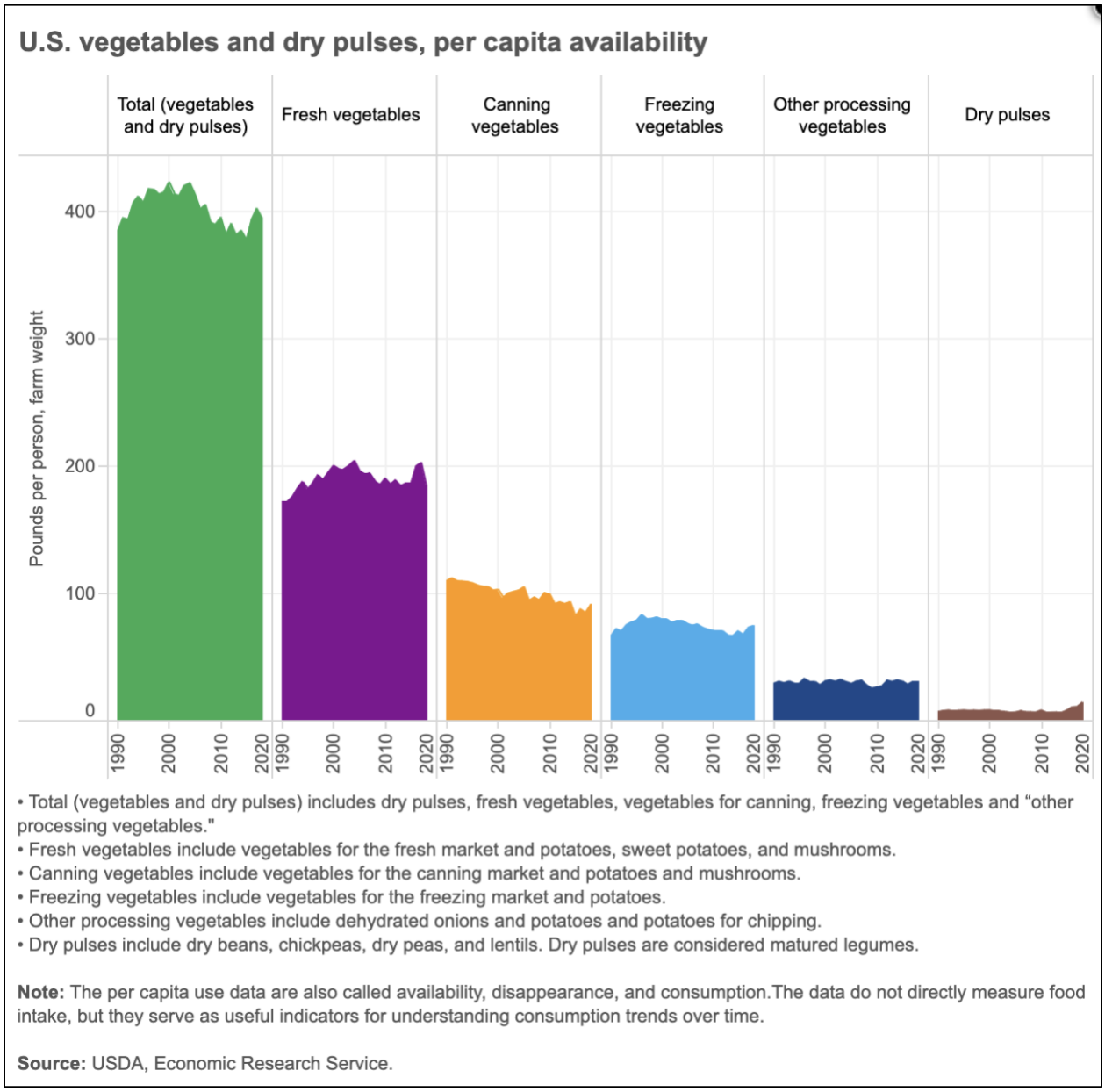


https://www.eia.gov/outlooks/aeo/pdf/aeo2019.pdf

COMMENTS:

- A way to combine two similar line graphs into a single figure that are able to share a legend.
- The legend prints the category levels in the same color as the lines in the figure therefore eliminating the need to include the line in the legend.
- A reference line is included and clearly labeled; it might be more visible if the line type was different, for example a dashed line.
- The range can be decreased from [0-25] to [0-20] which might help in seeing differences between the lines in the graph on the right-hand side.

Area Graph



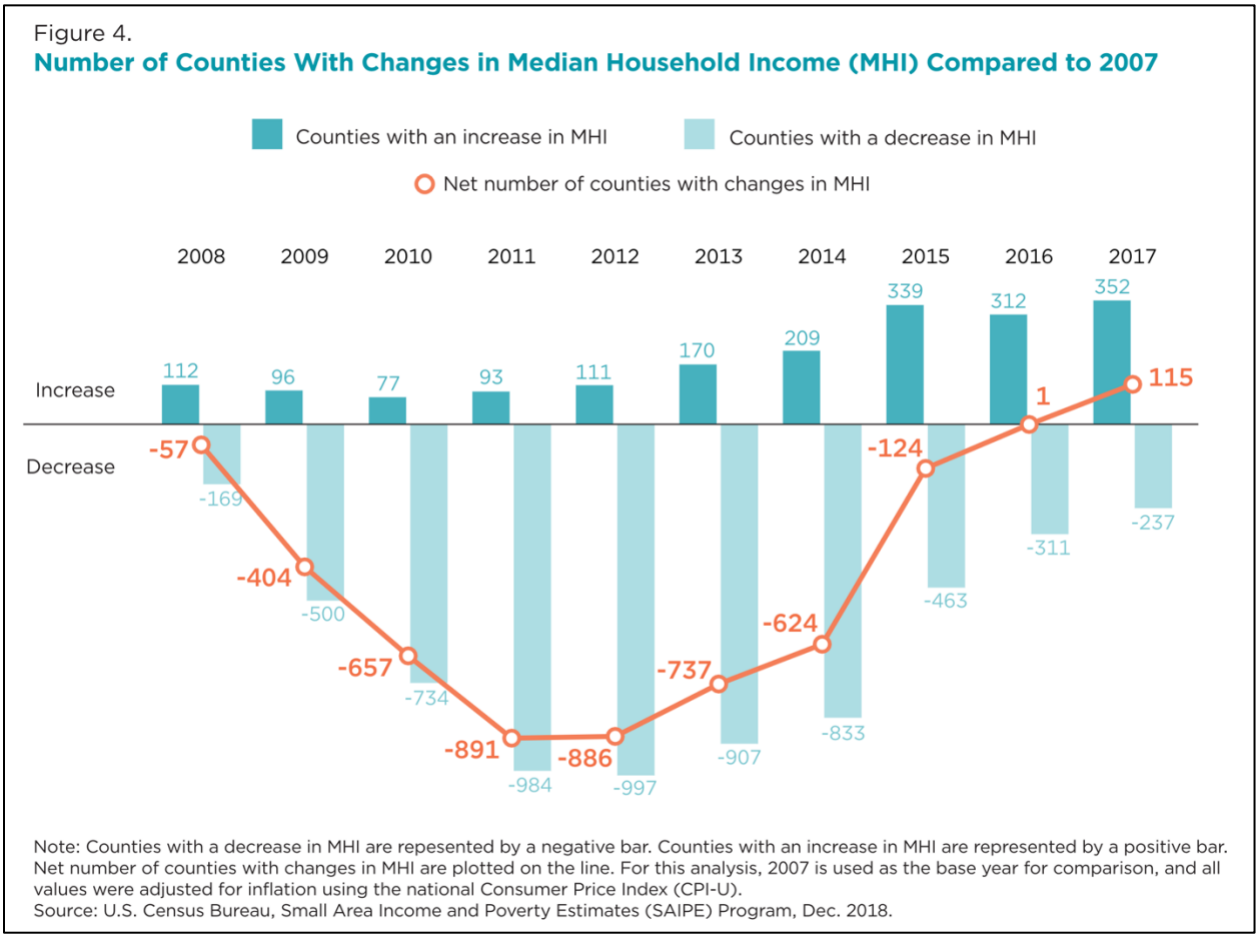
https://public.tableau.com/views/U_S_vegetablesanddrypulsespercapitaavailability/PerCapitaUse?amp&:display_count=no&amp&publish=yes&:embed=y&amp&:origin=viz_share_link&:toolbar=no&:showVizHome=no

COMMENTS:

- Rather than combining all levels in a single display using a line or area graph, each level of the category is displayed separately as an area graph, giving the impression of high magnitude of differences, but difficulty comparing trends over time.
- Since the bar on the far left is plotting reference values it could be plotted underneath each of the other five bars as a shade of grey – this might help with interpretation.

5. Graphs that Display a Difference

Multiple Visualizations: Divergent Bar Chart and Line Graph

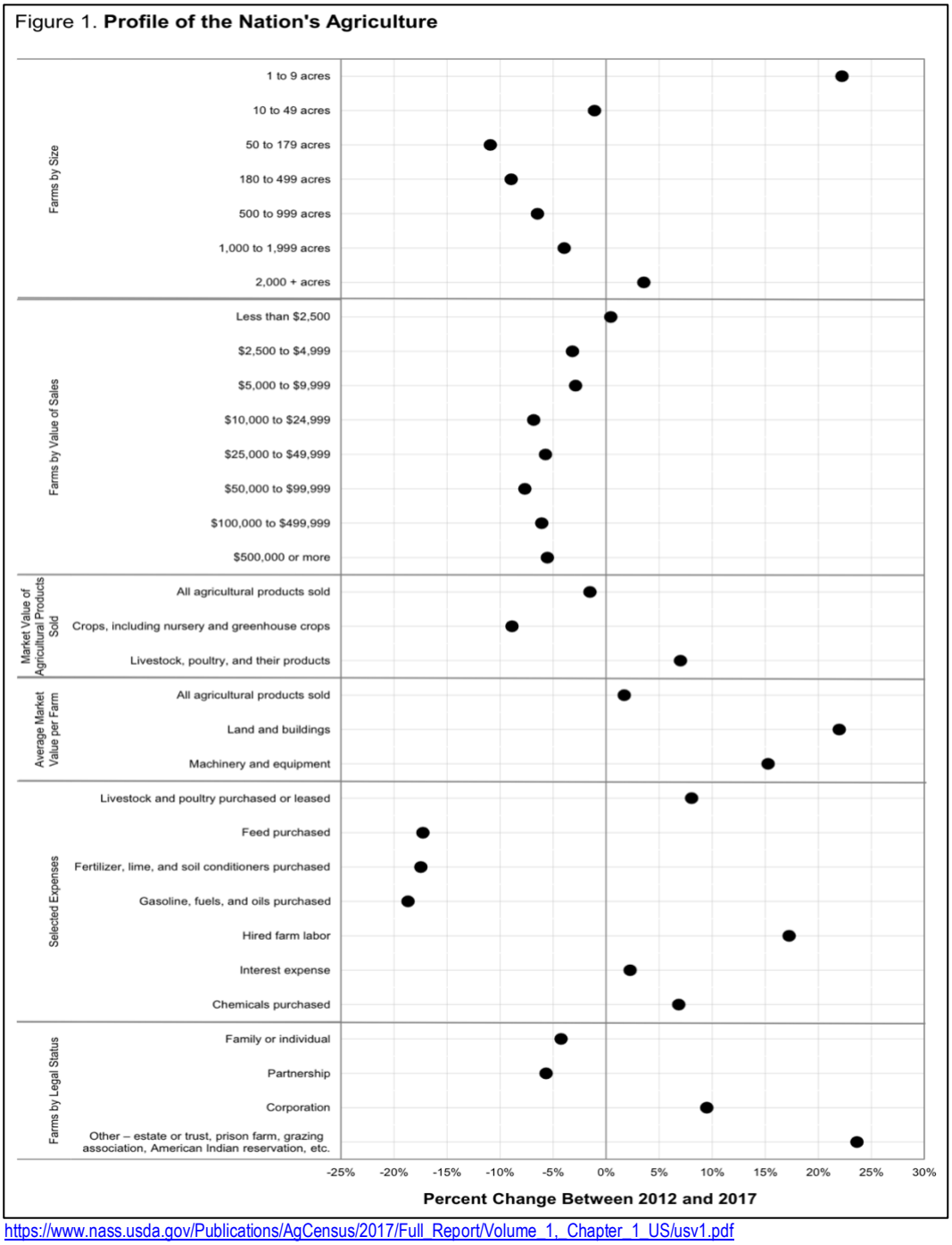


https://www.census.gov/content/dam/Census/library/publications/2018/demo/p30-04.pdf

COMMENTS:

- A way to display two quantitative variables over time along with their difference.

Scatter Plot

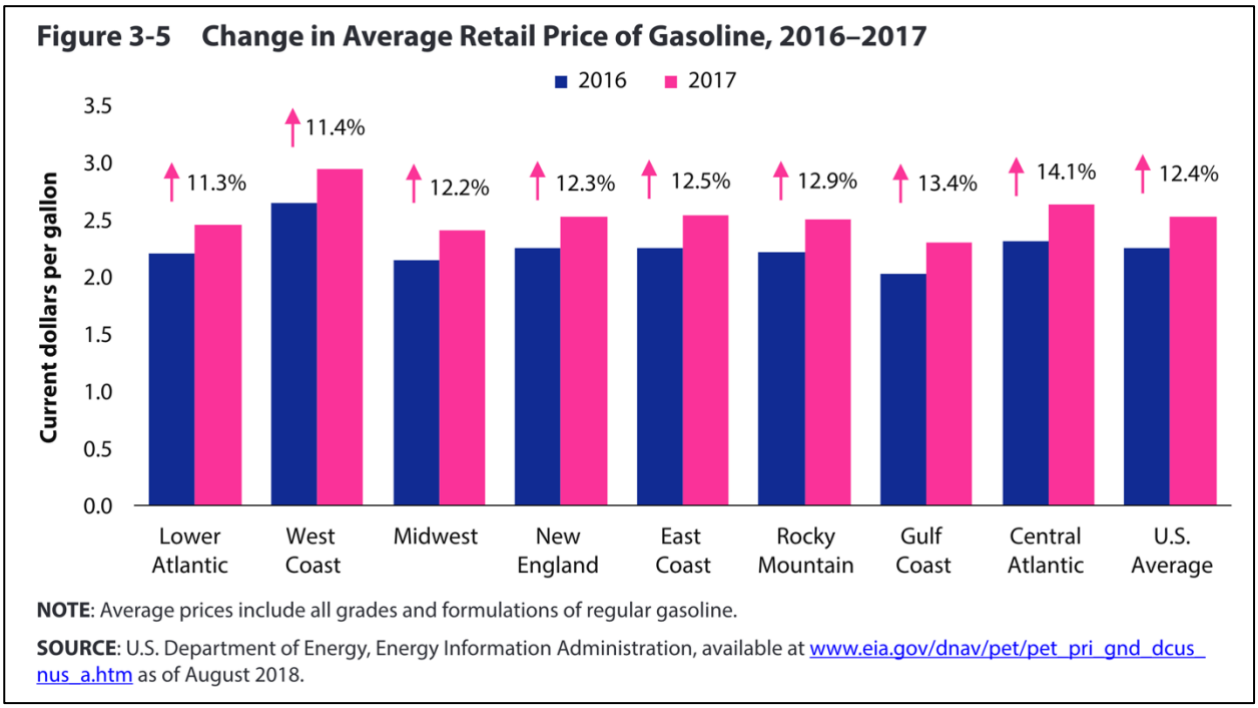


https://www.nass.usda.gov/Publications/AgCensus/2017/Full_Report/Volume_1,_Chapter_1_US/usv1.pdf

COMMENTS:

- A way to display differences among many similar categories using the same scale, percent change.
- Appropriate in cases where color is not an option.

Multi-set Bar Chart

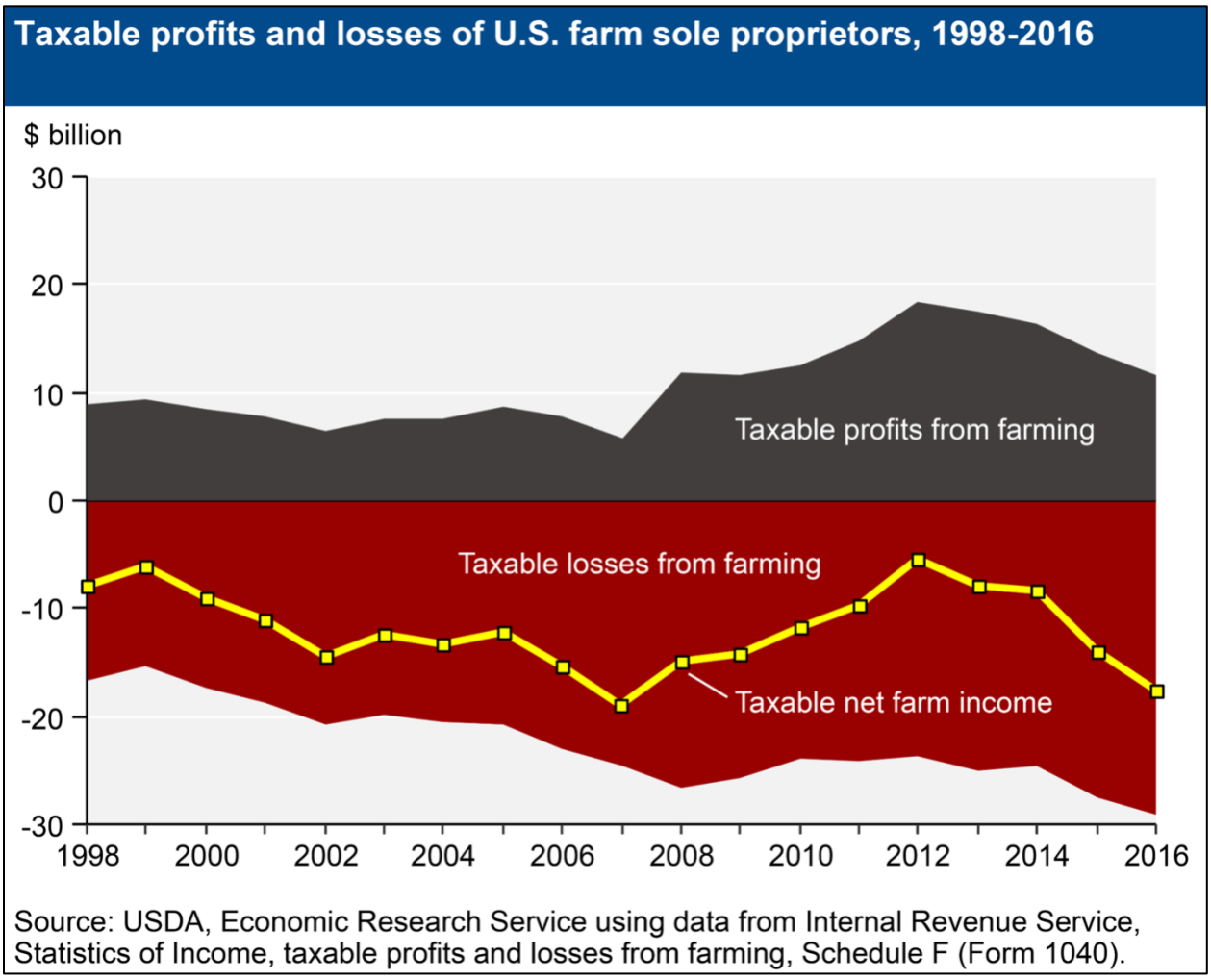


https://www.bts.gov/sites/bts.dot.gov/files/u796/TET%202018.pdf

COMMENTS:

- A relative percent difference is displayed using the value and an arrow; the two distributions used in calculating the relative percent difference are displayed using a multi-set bar chart organized by region with a reference bars at the far right.

Multiple Visualizations: Area Graph and Line Chart



https://www.ers.usda.gov/data-products/chart-gallery/gallery/chart-detail/?chartId=81953

COMMENTS:

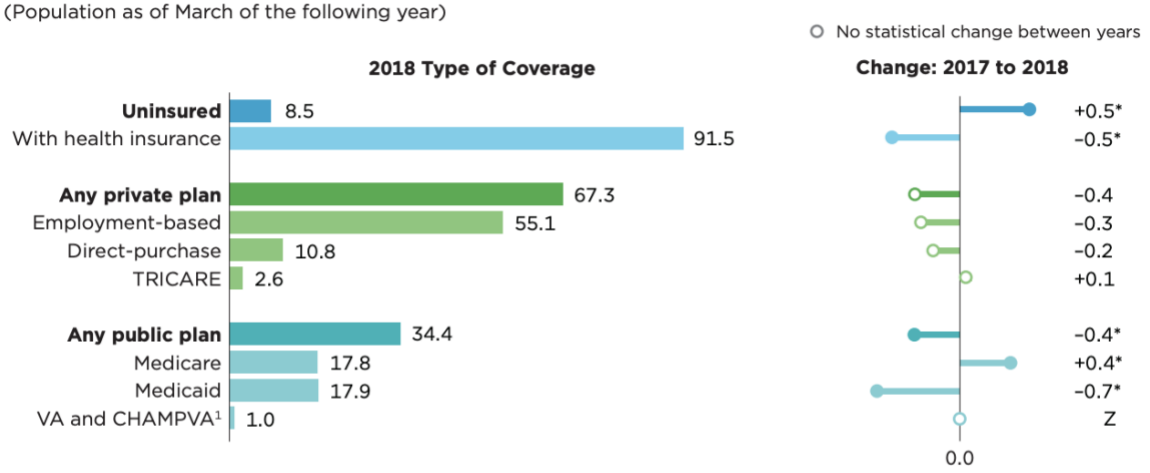
- A way to display a difference using area graphs for the two distributions and a line graph for the difference.

Multiple Visualizations: Multi-set Bar Chart and Lollipop Chart

Figure 1.

Percentage of People by Type of Health Insurance Coverage and Change From 2017 to 2018

(Population as of March of the following year)



Z Represents zero or rounds to zero.

[1] Includes CHAMPVA (Civilian Health Medical Program of the Department of Veterans Affairs), as well as care provided by the Department of Veterans Affairs and the military.

* Denotes a statistically significant change between 2017 and 2018 at the 90 percent confidence level.

Note: The estimates by type of coverage are not mutually exclusive; people can be covered by more than one type of health insurance during the year. For information on confidentiality protection, sampling error, nonsampling error, and definitions in the Current Population Survey, see <https://www2.census.gov/programs-surveys/cps/techdocs/cpsmar19.pdf>.

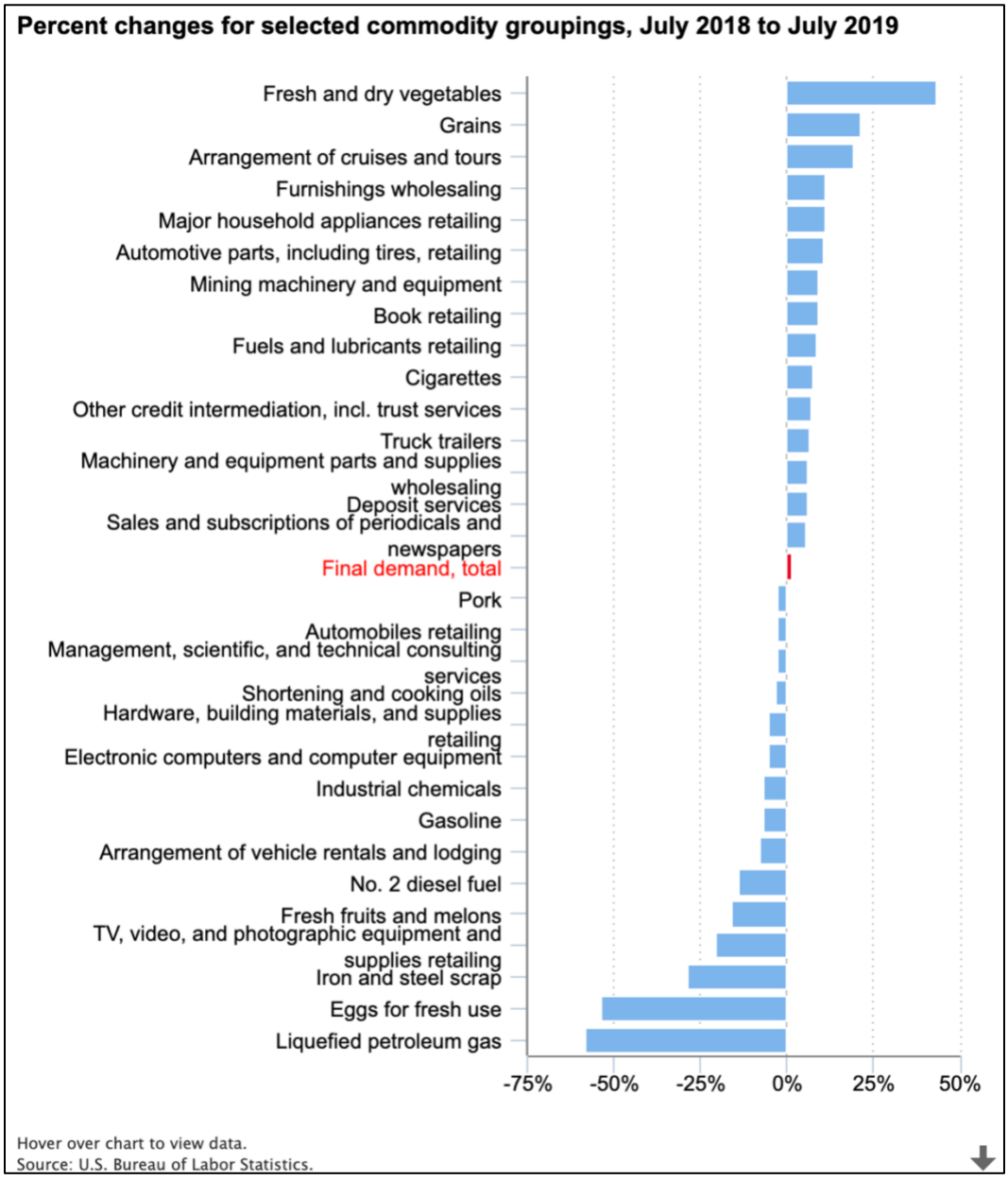
Source: U.S. Census Bureau, Current Population Survey, 2018 Annual Social and Economic Supplement Bridge File and 2019 Annual Social and Economic Supplement.

https://www.census.gov/library/visualizations/2019/demo/p60-267.html

COMMENTS:

- A multi-set bar chart is used to display the 2018 data (left) and the lollipop chart on the right side displays the change from the previous year (2017).
- The statistical difference between the years is displayed using different symbols in the lollipop chart; open circle for no statistical difference between 2017 and 2018 and a filled circle for a statistical difference.

Bar Chart



https://www.bls.gov/opub/ted/2019/producer-price-index-for-final-demand-increased-1-point-7-percent-for-the-year-ended-july-2019.htm

COMMENTS:

- A simple bar chart is used to display the percent difference, but by ordering the bars from top – largest positive percent difference, to bottom – largest negative percent difference. The ordering makes it easier to make comparisons.

Example of a diverging bar chart using NCSES data

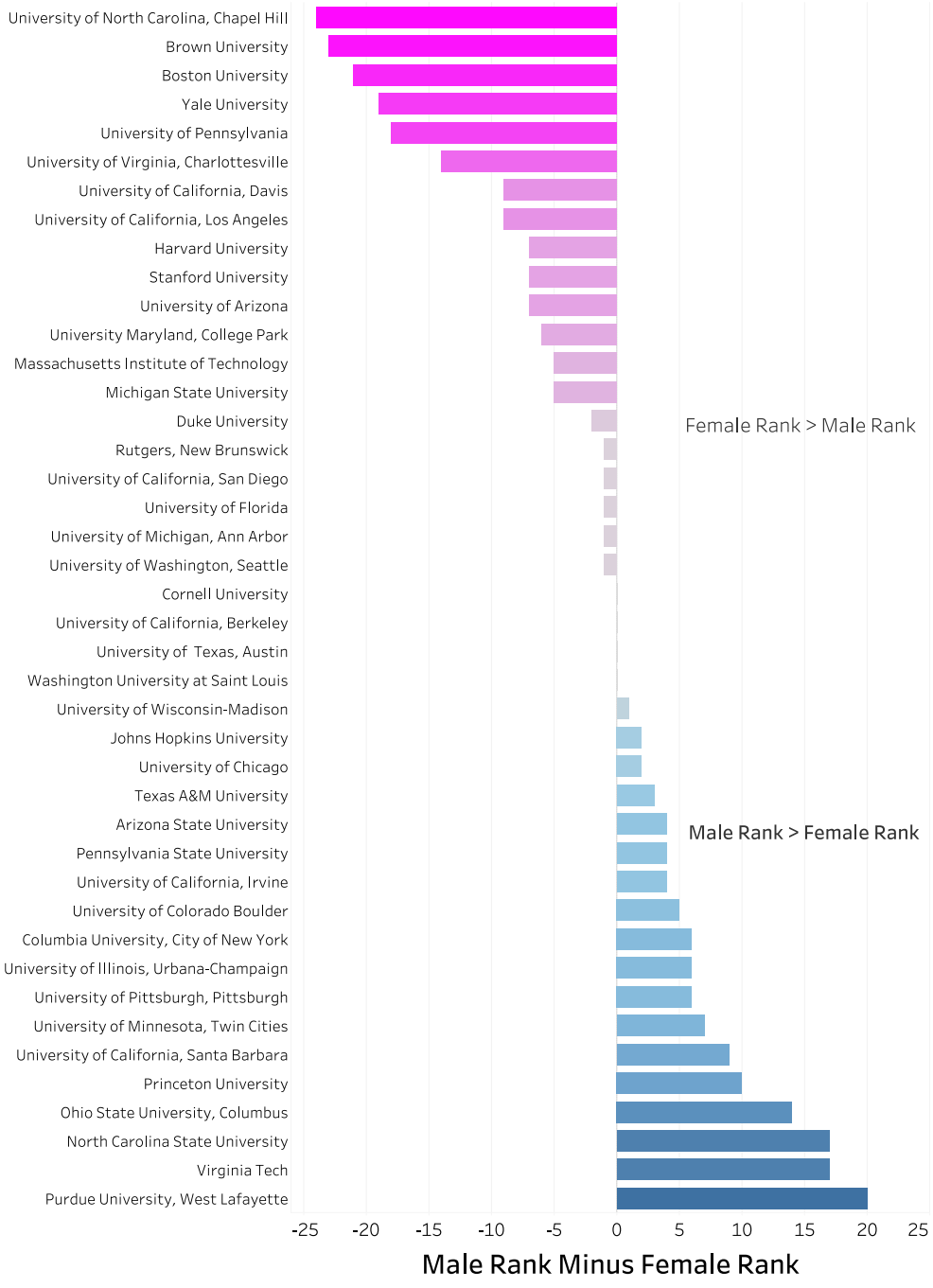
Women, Minorities, and Persons with Disabilities in Science and Engineering
SPECIAL REPORT | NSF 19-304 | March 08, 2019

TABLE 7-9
Top baccalaureate institutions of S&E doctorate recipients, by sex: 2013–17
(Number)

Academic institution	Doctorate recipients
Female, all institutions	78,345
Foreign or unknown institutions	28,829
Top 50 U.S. institutions	16,585
U. California, Berkeley	892
Cornell U.	637
U. Michigan, Ann Arbor	546
U. California, Los Angeles	542
Massachusetts Institute of Technology	510
U. Florida	495
U. Wisconsin-Madison	487
U. California, Davis	473
U. Texas, Austin	448
U. California, San Diego	428
U. Illinois, Urbana-Champaign	407
Pennsylvania State U., University Park and Hershey Medical Center	391
Harvard U.	381
U. Washington, Seattle	369
U. Maryland, College Park	348
Stanford U.	343
U. Virginia, Charlottesville	337
U. North Carolina, Chapel Hill	333
Michigan State U.	323
U. Pennsylvania	309
Texas A&M U., College Station and Health Science Center	303
U. Arizona	303
Brown U.	301
New York U.	300
Rutgers, State U. New Jersey, New Brunswick	298
U. Minnesota, Twin Cities	297
Boston U.	287
Yale U.	280
U. Chicago	270
Ohio State U., Columbus	256
Florida State U.	255
Purdue U., West Lafayette	255
U. Colorado Boulder	255
Duke U.	251
U. Georgia	247
Northwestern U.	242
Arizona State U.	240
Columbia U. in the City of New York	240
Virginia Polytechnic Institute and State U.	240
Princeton U.	238
U. California, Irvine	238
North Carolina State U.	235
U. California, Santa Barbara	231
U. Rochester	226
Johns Hopkins U.	221
Wellesley C.	220
Smith C.	217
U. California, Santa Cruz	216
Washington U., Saint Louis	214

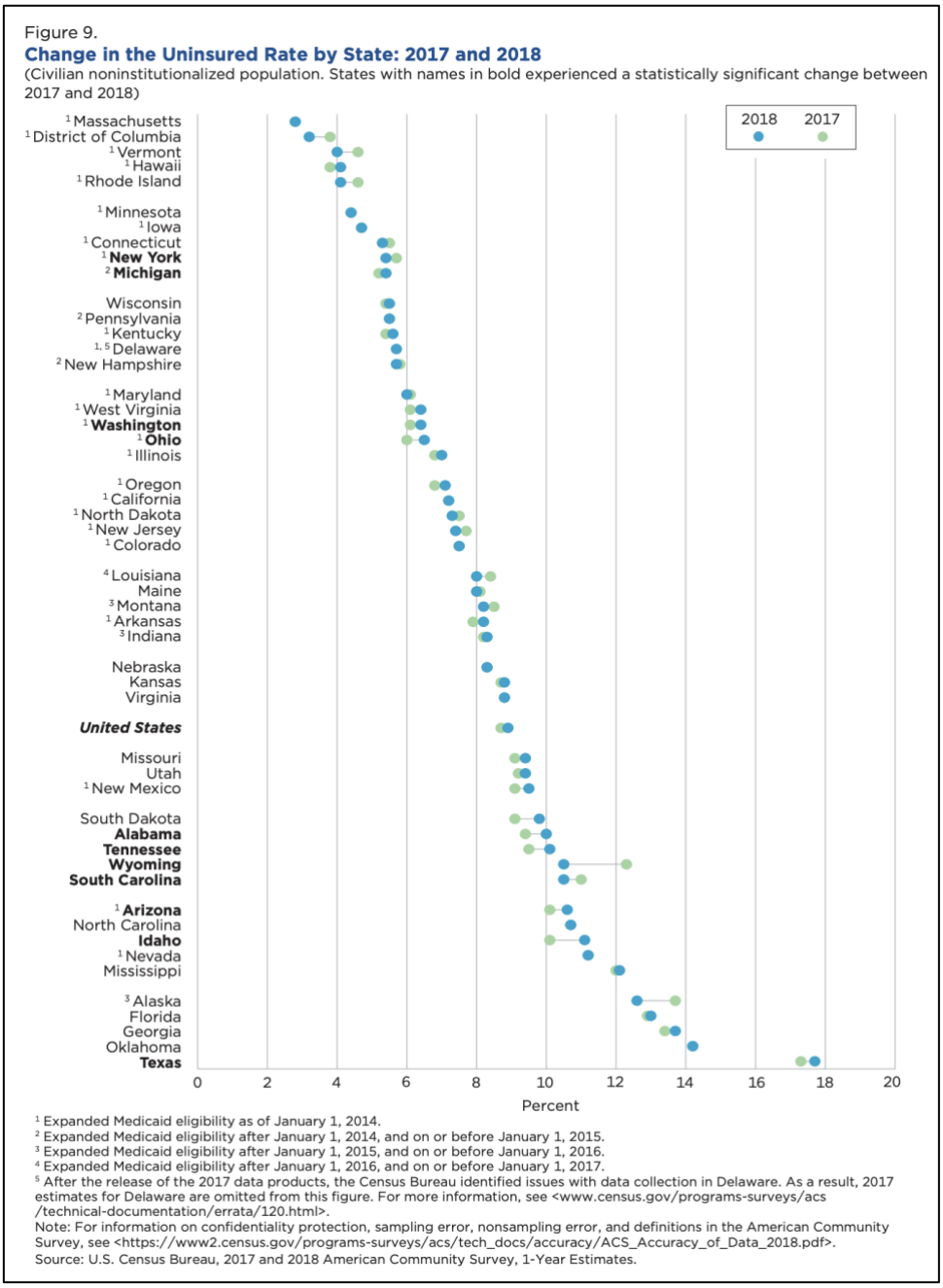
Only the first page of Table 7-9 is displayed, click on Data Tables to view the entire table.

Top Baccalaureate Institutions of S&E Doctorates: Difference in the Rank of Popularity Between Males and Females



Source: NCSES, NSF 19-304, March 2019
Includes the institutions in the Top 50 for both Males and Females who go on to earn a S&E doctorate. Institutions in the Top 50 for only one sex are omitted.

Dumbbell Dotplot (Cleveland Dot Plot)

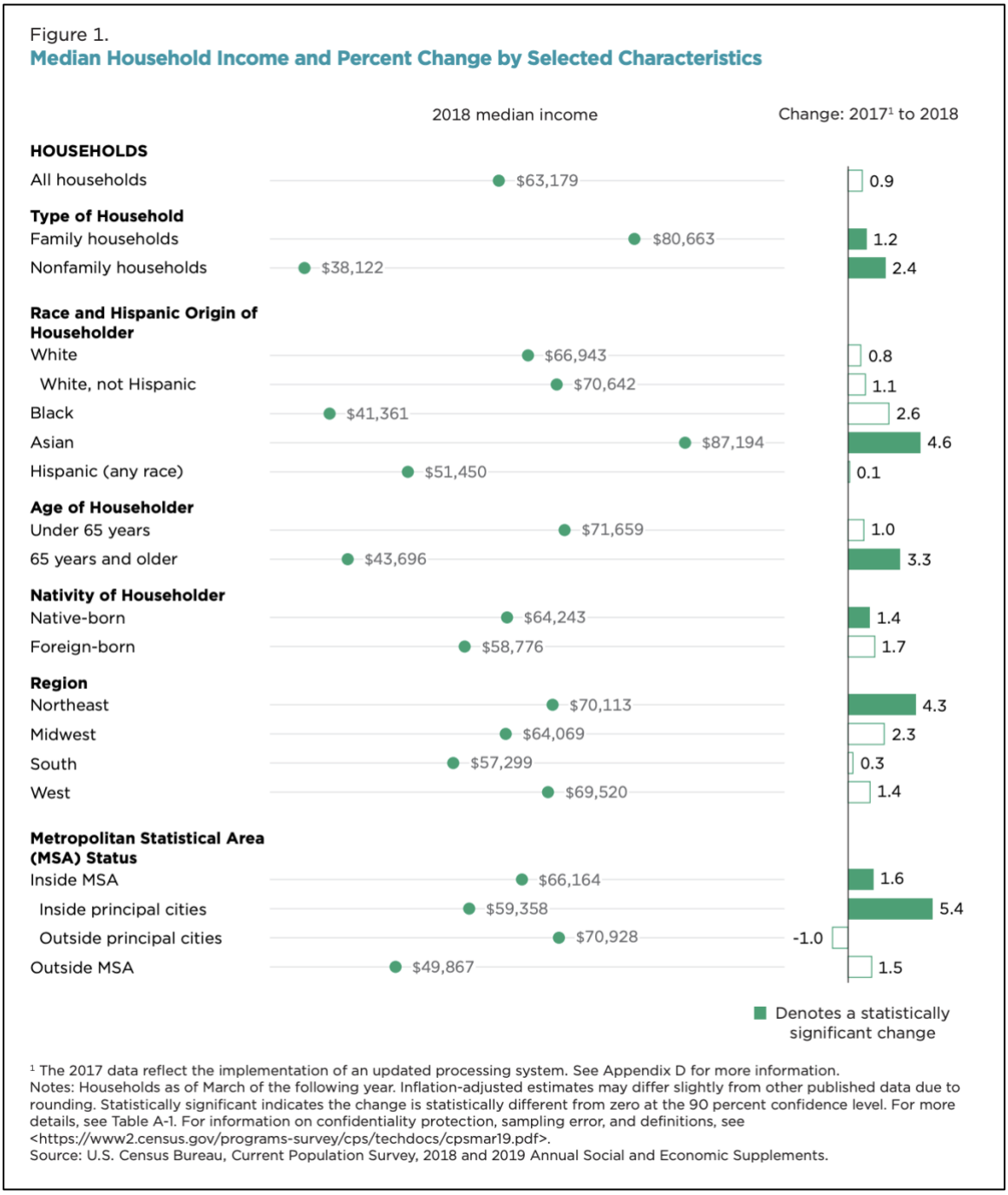


https://www.census.gov/content/dam/Census/library/visualizations/2019/demo/p60-267/Figure_9.pdf

COMMENTS

- Can be used in place of a stacked bar chart, it uses less ink and is easier to compare difference between values.
- A way to display a statistically significant change.

Multiple Visualizations: Bar Chart and Scatter Plot



https://www.census.gov/content/dam/Census/library/visualizations/2019/demo/p60-266/figure1.pdf

COMMENTS:

- Another way to display statistically significant differences using two plots in a single figure.

Example of a multiple visualizations using NCSES data

Science and Engineering Indicators 2018
NSB-2018-1 | 2018

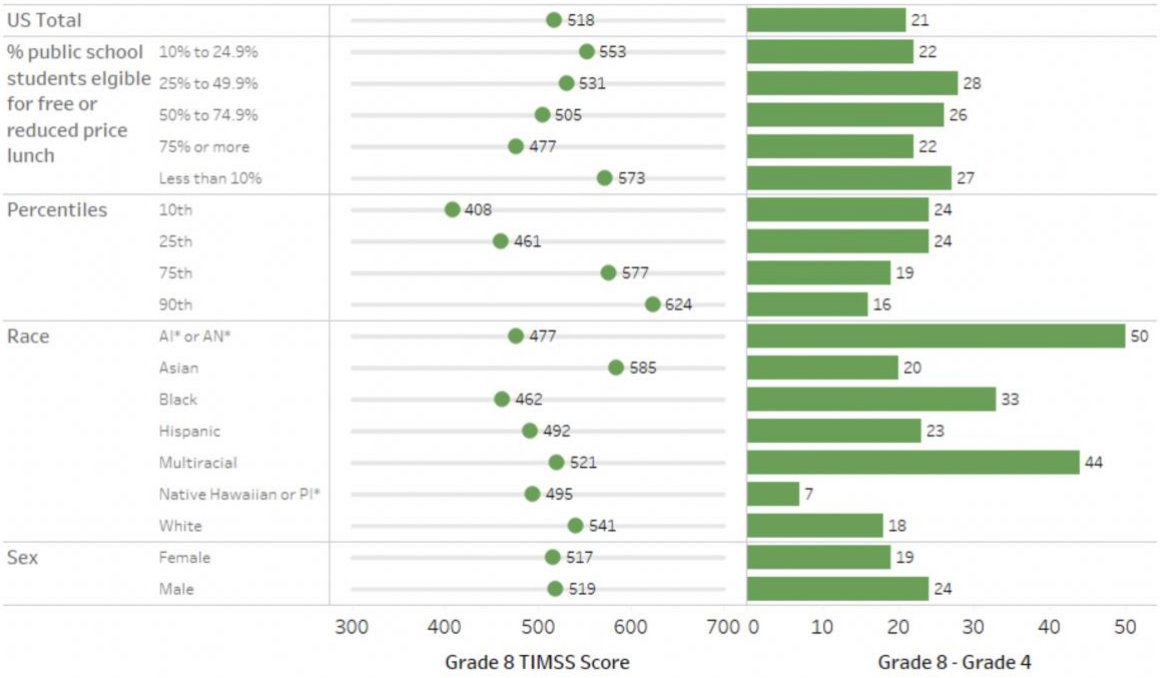
TABLE 1-6

Average TIMSS mathematics scores of U.S. students in grades 4 and 8, by selected student and school characteristics: 2015

(Average score)

Characteristic	Grade 4	Grade 8
U.S. total	539	518
Sex		
Male	543	519
Female	536	517
Race or ethnicity		
White	559	541
Black	495	462
Hispanic	515	492
Asian	605	585
Native Hawaiian or Pacific Islander	502	495
American Indian or Alaska Native	527	477
Multiracial	565	521
Percentage of public school students eligible for free or reduced-price lunch		
Less than 10%	600	573
10% to 24.9%	575	553
25% to 49.9%	559	531
50% to 74.9%	531	505
75% or more	499	477
Percentiles		
10th percentile	432	408
25th percentile	485	461
75th percentile	596	577
90th percentile	640	624

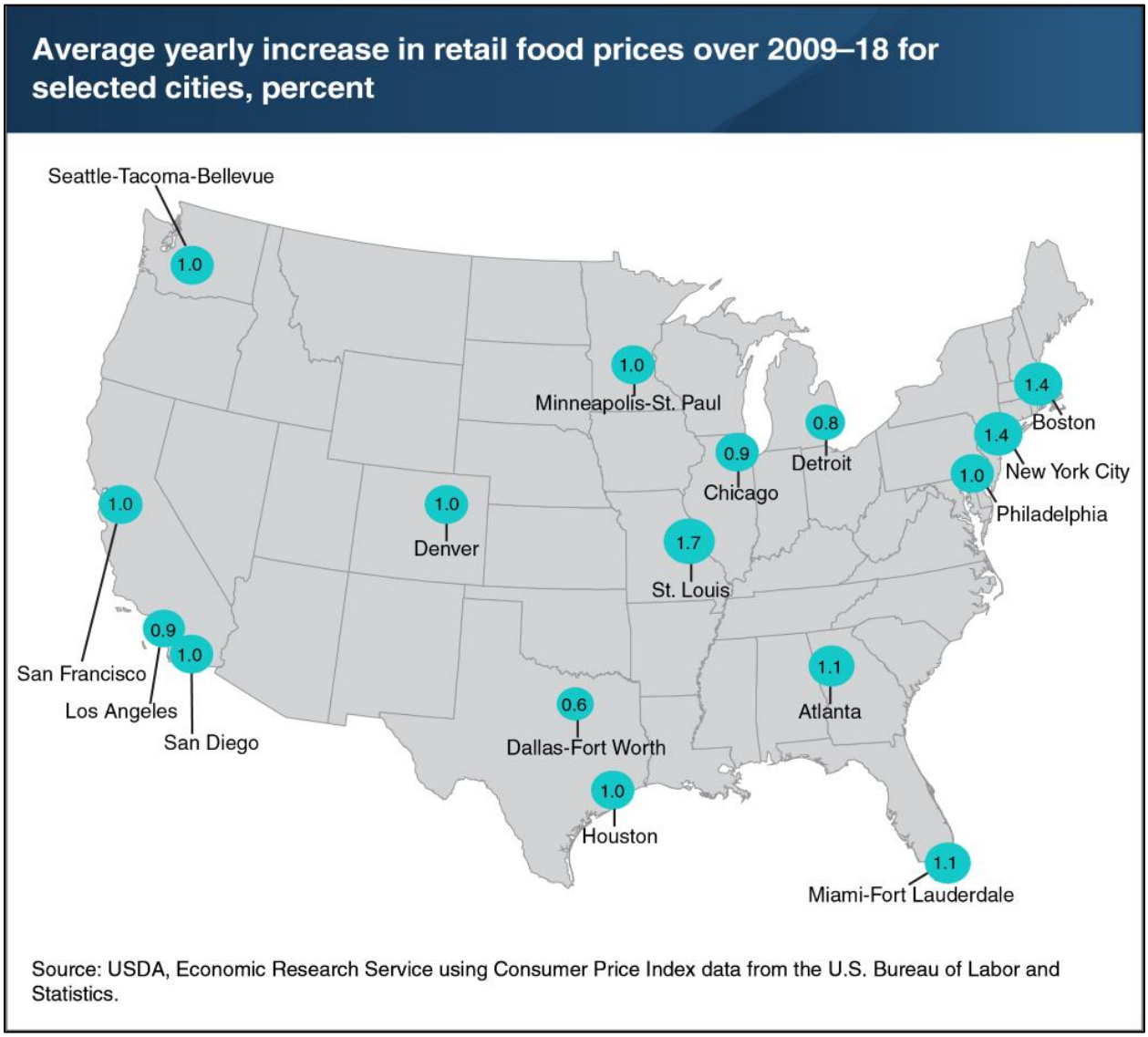
Average TIMSS mathematic scores of U.S. 8th grade students and the (8th – 4th) average differences by demographic variables and school characteristics: 2015



*AI=American Indian; AN=Alaska Native; PI=Pacific Islander

6. Geospatial Graphs

Dot Map

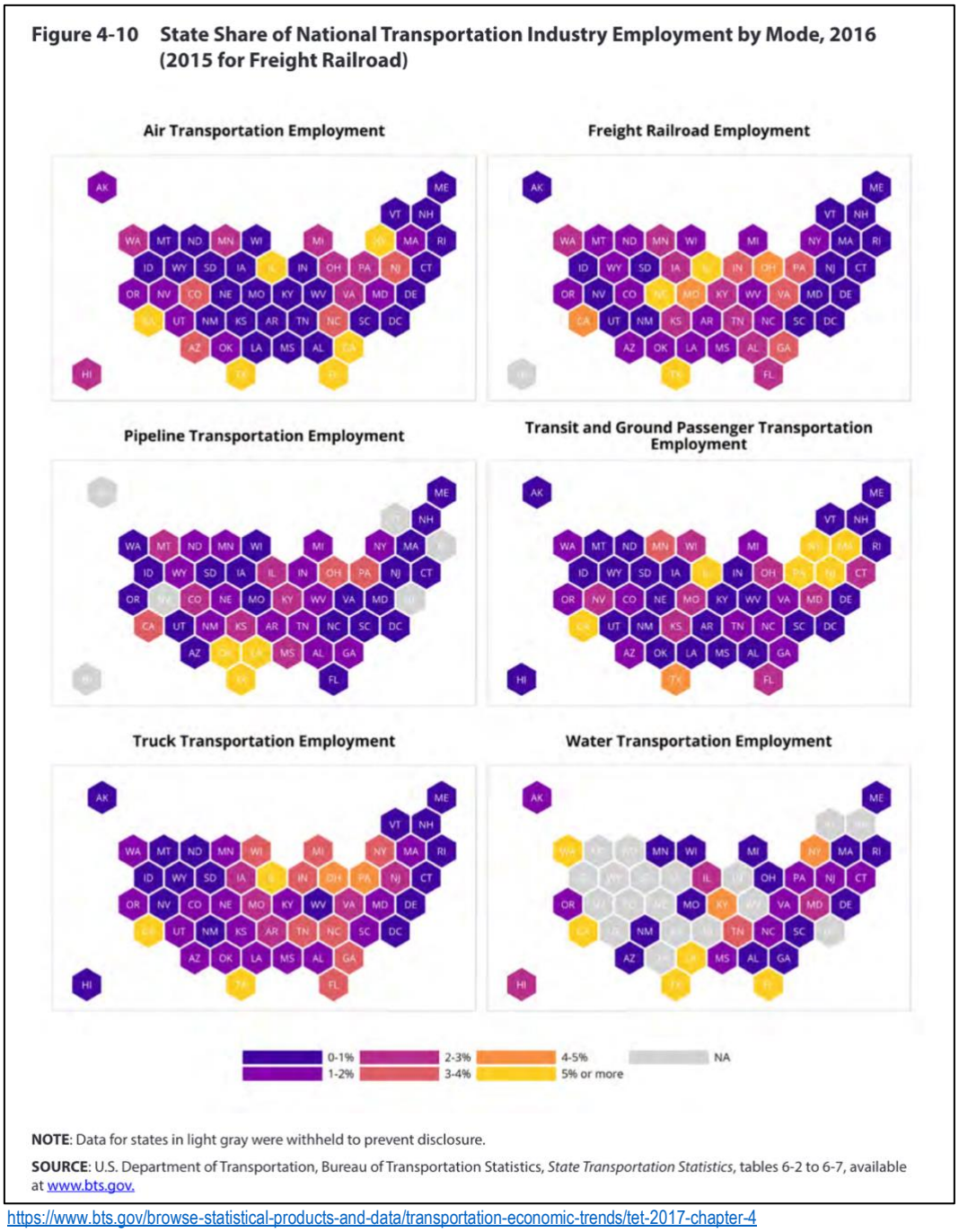


https://www.ers.usda.gov/data-products/chart-gallery/gallery/chart-detail/?chartId=92657

COMMENTS:

- Will only work if there are a small number of values that need to be displayed and they are not clustered together.
- It would be easier to see geographic trends if the color of the filled circles was a function of the value within the circle.

Hex-tile Chloropleth Maps



https://www.bts.gov/browse-statistical-products-and-data/transportation-economic-trends/tet-2017-chapter-4

COMMENTS:

- The visualization eliminates the size effect of states which makes it easier to make comparison within and between maps.
- Hex-file chloropleth maps are a product of Tableau.

Examples of a Hex-tile Chloropleth Maps using NCSES data

Science and Engineering Indicators 2018
NSB-2018-1 | January 15, 2018

TABLE 3-C

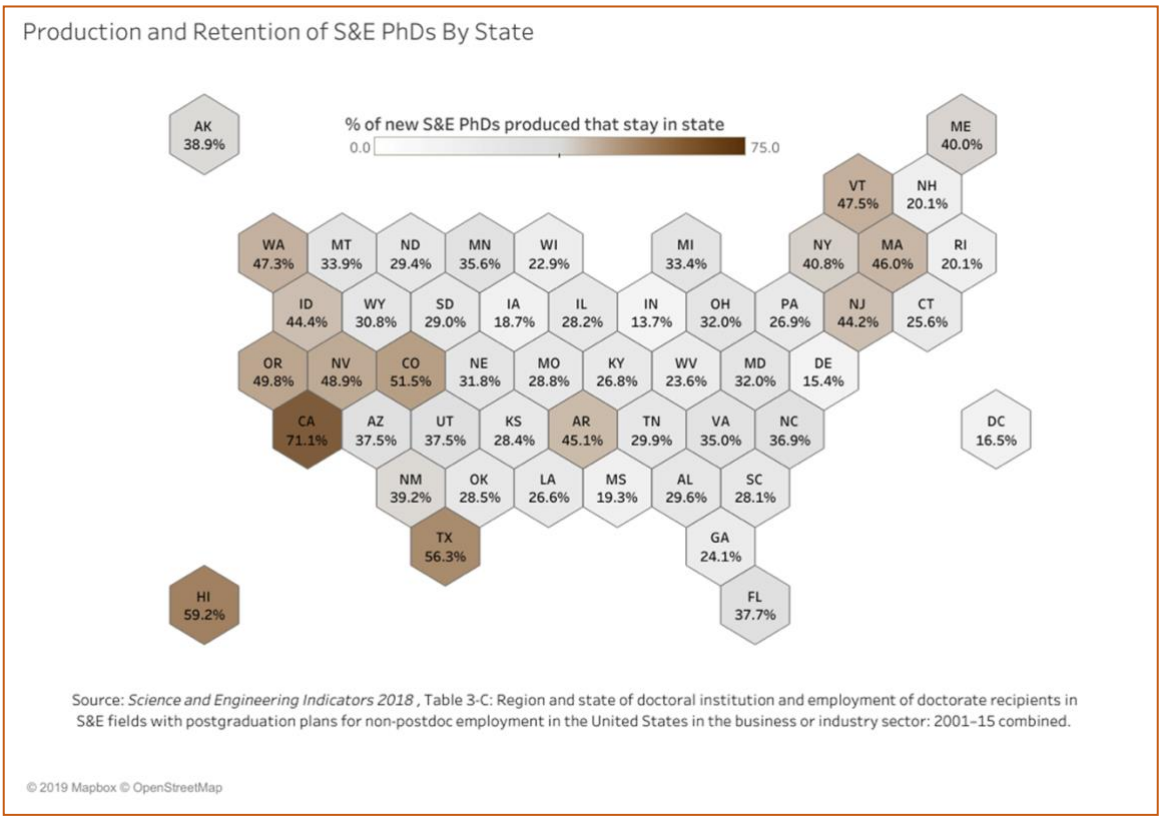
Region and state of doctoral institution and employment of doctorate recipients in S&E fields with postgraduation plans for non-postdoc employment in the United States in the business or industry sector: 2001–15 combined

(Number)

Region and state	PhDs trained in state or region	New PhDs working in state or region	Number of new PhDs produced that stay in state or region	Percent of new PhDs produced that stay in state or region
New England	4,566	4,762	2,212	48.4
Connecticut	669	859	171	25.6
Maine	40	57	16	40.0
Massachusetts	3,379	3,401	1,556	46.0
New Hampshire	159	195	32	20.1
Rhode Island	279	126	56	20.1
Vermont	40	124	19	47.5
Middle Atlantic	9,106	9,601	4,700	51.6
New Jersey	1,561	2,700	690	44.2
New York	4,273	4,741	1,744	40.8
Pennsylvania	3,272	2,160	879	26.9
East North Central	10,212	6,249	3,867	37.9
Illinois	3,291	2,149	929	28.2
Indiana	1,641	780	225	13.7
Michigan	2,100	1,502	701	33.4
Ohio	1,913	1,165	613	32.0
Wisconsin	1,267	653	290	22.9
West North Central	3,756	2,403	1,342	35.7
Iowa	761	274	142	18.7
Kansas	451	253	128	28.4
Minnesota	1,311	1,124	467	35.6
Missouri	820	527	236	28.8

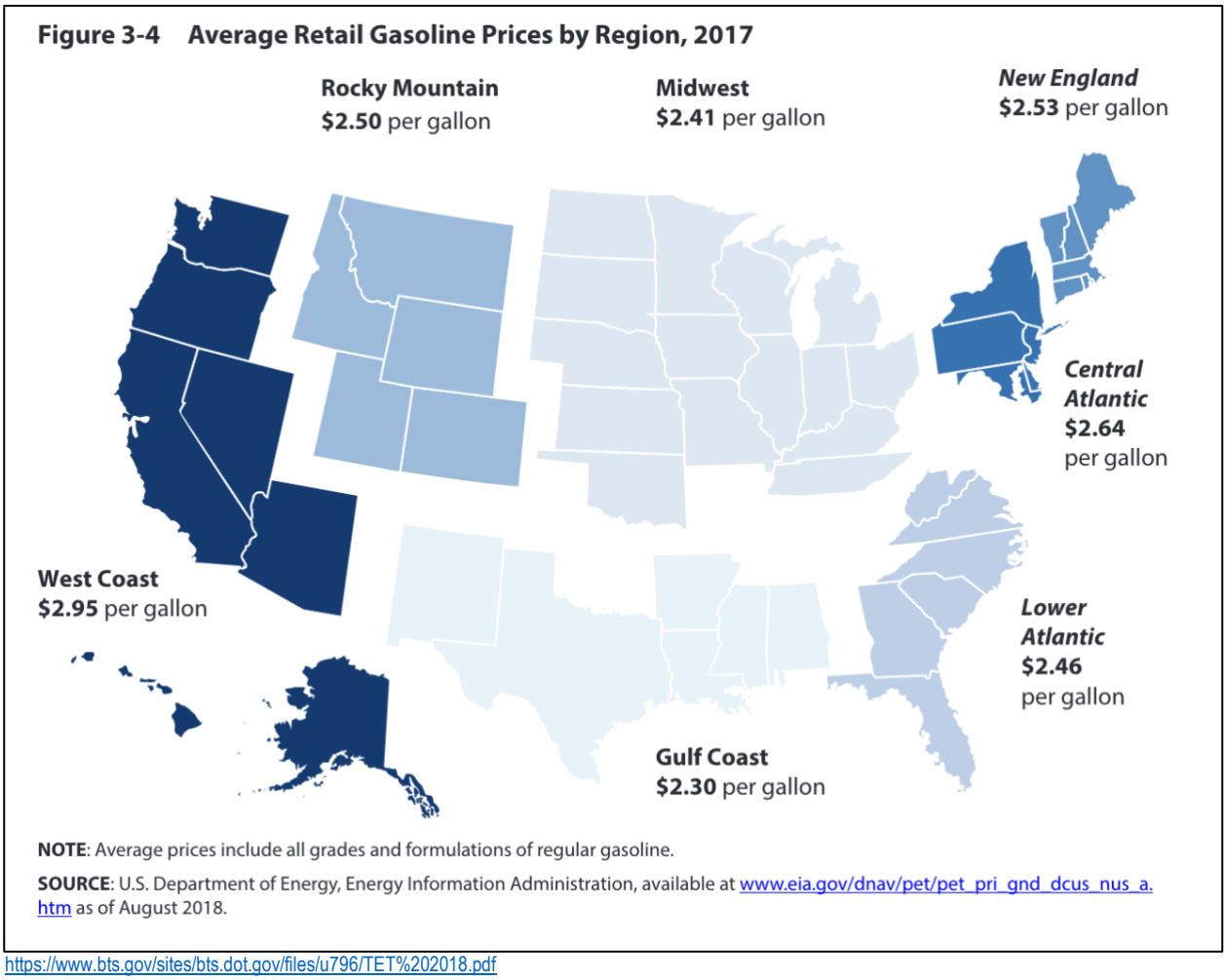
Only the first page of Table 3-C is displayed, see pages 49 & 50 of the report for the remaining data.

The chloropleth is created using a custom polygon in Tableau, downloaded from VizPainter by Joshua Milligan.



The map below uses a pie chart within each state to display the percentage of S&E Ph.Ds. produced in the state who have remained or left the state. The diameter of the pie chart is a function of the number of S&E Ph.Ds. produced by the state.

Chloropleth Map

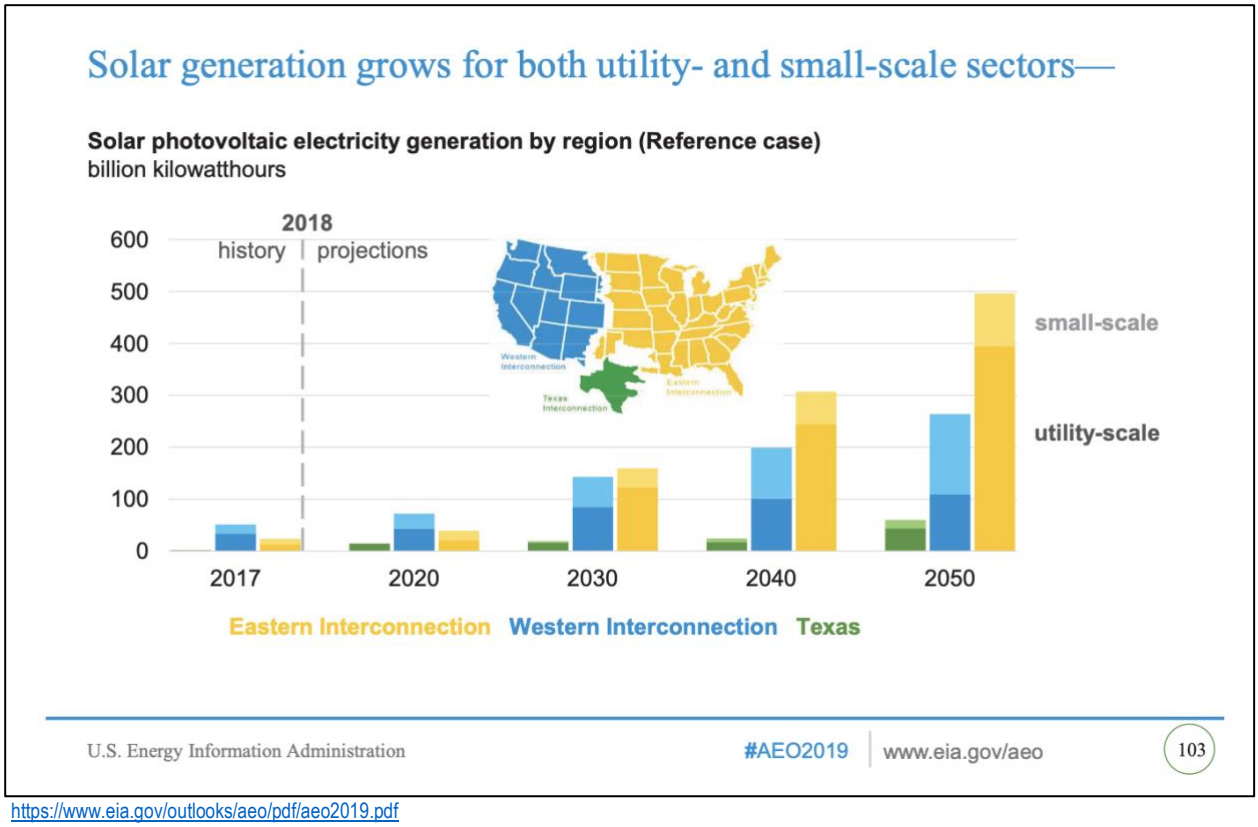


https://www.bts.gov/sites/bts.dot.gov/files/u796/TET%202018.pdf

COMMENTS:

- A way to differentiate between regions that is easier to interpret than outlining the states within each region.
- This map may be difficult to interpret for someone who is colorblind. As a rule-of-thumb one should use a different color for each of the seven regions rather than different shades.

Multiple Visualizations: Multi-set Bar Chart and Chloropleth Map

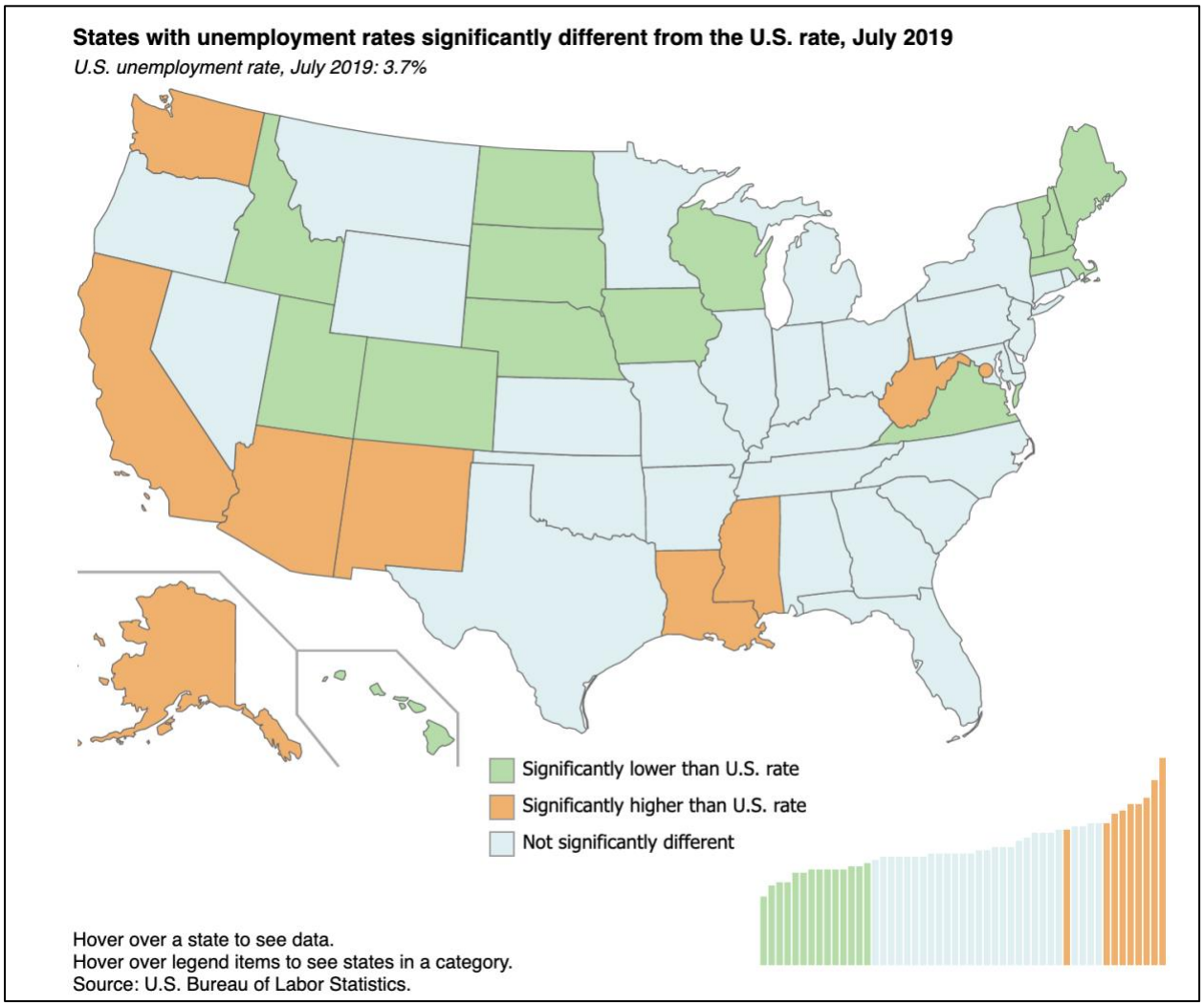


https://www.eia.gov/outlooks/aeo/pdf/aeo2019.pdf

COMMENTS:

- A way to include a region categorical variable with two other variables, one that is longitudinal and the other categorical with two levels.

Multiple Visualizations: Bar Chart and Chloropleth Map



https://www.bls.gov/opub/ted/2019/unemployment-rate-lowest-in-vermont-highest-in-alaska-july-2019.htm

COMMENTS:

- The map displays state unemployment rates categorized into three levels based on the statistical difference from the U.S. unemployment rate. The bar chart in the bottom right corner displays the actual rates ordered from lowest to highest. If this was used in a static document one could label each bar with the state abbreviation and add a vertical axis.

7. Other

Heatmap Table

State comparisons with the national average

How did the household budgets in California, Texas and New York compare with the national average? Looking at household budgets, together with the national average, gives us a better view of where household spending is the highest and lowest. Table 5 shows household budget shares for the three states, compared with the national budget share. Each state's column represents that state's budget share. A blue box indicates spending share below the national share and a red box indicates spending share above the national share.

Table 1. State-level budget shares compared with the national budget share, 2017

Expenditure category	National share	California share	Texas share	New York share
Average annual expenditure	$60,060	$64,333	$54,908	$59,323
Food	13	15	15	15
Food at home	7	10	10	10
Food away from home	6	5	5	5
Housing	33	37	33	39
Apparel and services	3	2	2	2
Transportation	16	15	20	14
Healthcare	8	6	7	6
Entertainment	5	4	4	4
Education	2	3	3	3
Cash contributions	3	2	3	2
Personal insurance and pensions	11	13	11	13
Pensions and Social Security	11	12	10	12
All other expenditures	5	3	3	3

Source: U.S. Bureau of Labor Statistics.

https://www.bls.gov/opub/btn/volume-8/pdf/consumer-spending-ca-tx-ny.pdf

COMMENTS:

- A way to aid in the interpretation of a table by adding color based on a reference value, red – above and blue – below, in this case the reference value is displayed in the first column.

Example of a Heatmap Table using NCSES data

Science and Engineering Indicators 2020: Science and Engineering Labor Force 2020
NSB-2019-8 | 2019

TABLE 3-A

Bureau of Labor Statistics projections of employment and job openings in S&E and other selected occupations: 2016–26

(Thousands)

Occupation	BLS National Employment Matrix 2016 estimate	BLS projected 2026 employment	Occupational openings, 2016–26, annual average	10-year growth in total employment (%)
Total, all occupations	156,063.8	167,582.3	18,742.0	7.4
All S&E	6,952.6	7,825.3	591.5	12.6
Computer and mathematical scientists (excluding computer programmers, including logisticians)	4,248.7	4,882.3	364.7	14.9
Engineers, including ship engineers and sales engineers	1,765.8	1,911.0	136.1	8.2
Life scientists	325.4	358.0	32.4	10.0
Physical scientists	278.2	305.3	28.0	9.7
Social and related scientists (excluding historians)	334.5	368.7	30.3	10.2
S&E-related occupations				
S&E managers	956.6	1,088.4	88.0	13.8
S&E technicians and technologists, except computer programmers	1,125.2	1,203.7	113.5	7.0
Computer programmers	294.9	273.6	15.5	-7.2
Health care practitioners and technicians	8,751.5	10,088.1	625.1	15.3
Selected other occupations				
Lawyers	792.5	857.5	40.7	8.2
Postsecondary teachers	1,871.4	2,108.3	172.4	12.7

BLS = Bureau of Labor Statistics.

Note(s)

Estimates of current and projected employment for 2016–26 are from BLS's National Employment Matrix; data in the matrix are from the Occupational Employment Statistics (OES) Survey and the Current Population Survey (CPS). Together, these sources cover paid workers, self-employed workers, and unpaid family workers in all industries, agriculture, and private households. Because data are derived from multiple sources, they can often differ from employment data provided by the OES Survey, CPS, or other employment surveys alone. BLS does not make projections for S&E occupations as a group nor does it do so for some of the S&E and S&E-related occupational categories as defined by the National Science Foundation (NSF); numbers in the table are based on the sum of BLS projections for occupations that the NSF includes in the respective categories.

Source(s)

National Center for Science and Engineering Statistics, National Science Foundation, special tabulations (2018) of the 2016–26 BLS Employment Projections. See Table S3-1.

Science and Engineering Indicators

Bureau of Labor Statistics projections of S&E employment and job openings and other selected occupations: 2016-2016

	Occupations	BLS National Employment Matrix 2016 Estimates	BLS Projected Employment 2026 Estimates	Ave. Annual Employment Openings 2016-2026	10-YR % Growth in Total Employment
All S&E	Computer & Mathematics	4,2448.7	4,882.3	364.7	**14.9**
	Social & Related Scientists	334.5	368.7	30.3	**10.2**
	Life Scientists	325.4	358.0	32.4	**10.0**
	Physical Scientists	278.2	305.3	28.0	**9.7**
	Engineers	1,765.8	1,911.0	136.1	**8.2**
Related S&E	Healthcare Practitioners & Technicians	8,751.5	10,0881	625.1	**15.3**
	S&E Managers	956.6	1,088.4	88.0	**13.8**
	S&E Technicians & Technologists	1,125.2	1,203.7	113.5	**7.0**
	Computer Programmers	294.9	273.6	15.5	**-7.2**
Other	Postsecondary Teachers	1,871.4	2,108.3	172.4	**12.7**
	Lawyers	792.5	857.5	40.7	**8.2**
All S&E		6,952.6	7,825.3	591.5	**12.6**
Total		156,063.8	167,582.3	18,742.0	**7.4**

10-YR % Growth Legend					
-10 <	-10 to -5 <	-5 to -1 <	> 1 to 5	> 5 to 10	> 10

The data on page 46 are displayed as a heatmap where the color of a row is a function of the last column, 10-YR % Growth in Total Employment. Within the categories of All S&E, Related S&E, and Other, the rows are ordered by the last column from largest to smallest.

Sparklines

Table 3-1 Detailed Producer Price Indexes by Transportation Modes, 2004–2016 (2004 = 100)

Mode	2004	2017	2004-2017
Air transportation (NAICS 481)[1]	100.0	135.0	
Scheduled air transportation (NAICS 4811)[2]	100.0	135.4	
Scheduled freight air transportation (NAICS 481112)	100.0	150.1	
Nonscheduled air transportation (NAICS 4812)[3]	100.0	140.5	
Rail transportation (NAICS 482)[3]	100.0	160.3	
Line -haul railroads (NAICS 482111)[4]	100.0	160.3	
Water transportation (NAICS 483)	100.0	128.1	
Deep sea freight transportation (NAICS 483111)[5]	100.0	116.3	
Coastal and great lakes freight transportation (NAICS 483113)	100.0	147.6	
Inland water freight transportation (NAICS 483211)[6]	100.0	159.0	
Truck transportation (NAICS 484)	100.0	130.4	
General freight trucking (NAICS 4841)	100.0	132.7	
General freight trucking, local (NAICS 48411)	100.0	123.6	
General freight trucking, long distance (NAICS 48412)	100.0	134.7	
Specialized freight trucking (NAICS 4842)	100.0	125.2	
Used household and office goods moving (NAICS 48421)	100.0	124.1	
Specialized freight (except used goods) trucking, local (NAICS 48422)	100.0	129.8	
Specialized freight (except used goods) trucking, long distance (NAICS 48423)	100.0	120.8	
Pipeline transportation (NAICS 486)	NA	NA	
Pipeline transportation of crude oil (NAICS 4861)	100.0	224.1	
Other pipeline transportation (NAICS 4869)	100.0	170.8	
Support activities for transportation (NAICS 488)	100.0	120.0	
Support activities for water transportation (NAICS 4883)	100.0	137.6	
Postal service (NAICS 491)[2]	100.0	140.1	
Couriers and messengers (NAICS 492)	100.0	198.3	

NOTES: Blue dots on the sparkline charts indicate high values; red dots indicate low values. Transportation Mode defined by the North American Industry Classification System (NAICS). Indexes rebased, year 2004 = 100.

SOURCE: U.S. Department of Labor, Bureau of Labor Statistics, Producer Price Index, available at www.bls.gov/ppi as of June 2018.

https://www.bts.gov/sites/bts.dot.gov/files/u796/TET%202018.pdf

COMMENTS:

- The term sparkline was introduced by Edward Tufte for "small, high resolution graphics embedded in a context of words, numbers, images". Tufte described sparklines as "data-intense, design-simple, word-sized graphics".
- Tufte gave a workshop at the Department of Transportation.
- Excel employees filed a patent application for the implementation of sparklines.
- In the above figure, a box or horizontal reference line would aid the interpretation of the sparklines.