# PERCENTAGE GRAPHS IN USA TODAY SNAPSHOTS ONLINE

Milo Schield, Augsburg College

Department of Business Administration. Minneapolis, MN 55454

**Abstract**: This paper reviews 229 graphs presented in USA Today Snapshots online. About 70% of these graphs involve percentages and about 70% of these percentage graphs are bar graphs. Bars are wholes in 10% of these percentage bar graphs. Problems involving the part-whole status of the bars in percentage bar charts are analyzed. Rules are proposed to determine if the bars are wholes but are found to be highly dependent on the context. Some student feedback is included. More work is needed to justify part-whole choices based on content. Recommendations are offered.

## 1. INTRODUCTION

The 2006 GAISE College report recommended that "introductory courses in statistics should, as much as possible, strive to emphasize statistical literacy and develop statistical thinking." The authors defined statistical literacy as "understanding the basic language of statistics (e.g., knowing what statistical terms and symbols mean and being able to read statistical graphs), and understanding some fundamental ideas of statistics." The college report suggested that teachers assess statistical literacy by students "interpreting or critiquing articles in the news and graphs in media."

In accordance with the GAISE college guideline, this paper analyzes graphs from USA Today: the national newspaper with the largest circulation (over 2.5 million). USA Today features online graphs for math education<sup>1</sup> for careers<sup>2</sup> and for general interest.<sup>3</sup> The 229 graphs analyzed herein were obtained from the general interest "Snapshots" and are almost a census of graphs displayed between 11/2004 and 11/2006.

Graphs show the relationship between variables. Often these graphs involve percentages – a key element in statistical literacy since they take into account the size of a group. The 2002 W. M. Keck Statistical Literacy survey identified difficulties in describing and comparing rates and percentages in tables and graphs. See Schield (2004, 2000, 2001 and 2006)

## 2. GRAPHS IN USA TODAY

The graphs obtained from USA Today Snapshots online can be classified by their content as dealing with percentages, rates, measures and counts. These graphs can be classified by their form as pie charts<sup>4</sup>, bar charts and other. Figure 1 shows a pie chart and a bar chart.

Pie charts show the parts (slices) of a whole (pie). The bars in a bar chart can be parts or wholes.

#### Figure 1. Sample pie and bar charts



Table 1 classifies the USA Today Snapshot graphs

|--|

	ALL	Bar	Pie	Other
ALL	229	170	46	13
Percentages	164	116	46	2
Totals & Measures	57	46		11
Ratios & % Change	8	8		

Of these graphs, about 70% involve percentages. Of the 164 graphs involving part-whole percentages, about 70% use bar charts while the rest use pie charts.

## 3. GRAPHS OF COUNTS AND MEASURES

About 30% of these graphs involve totals, measures, non-percentage ratios or percentage changes.

## Figure 2. Graphs of Totals or Measures



**Figure 3. Graphs of Ratios or Percent Changes** 



Percentage change may get confused with part-whole percentages. All subsequent charts involve part-whole percentages so that qualification is often omitted.

## 4. PIE CHARTS

About 30% of these USA Today Snapshot online percentage graphs are pie charts (circle graphs). The slices

<sup>&</sup>lt;sup>1</sup> <u>www.usatoday.com/educate/mathtoday/index.htm</u>

<sup>&</sup>lt;sup>2</sup> www.usatodaysecure.com/snapshots/

<sup>&</sup>lt;sup>3</sup> <u>www.usatoday.com/news/snapshot.htm</u>

<sup>&</sup>lt;sup>4</sup> 'Chart' (diagram, table or illustration) and 'graph' (diagram representing values of a variable) are used interchangeably.

are parts that are exclusive and exhaustive so the percentages should add to 100%.

Groups that are naturally exclusive and exhaustive may involve two categories such as Yes/No.

## Figure 4. Pie Chart, Two Valued



In addition to the choices that are useful, there may be other responses that may not as valuable such as 'not sure,' 'no answer,' 'don't know,' and 'maybe.'<sup>5</sup>

## Figure 5. Pie Chart, Multi-Valued



Groups that are exclusive and exhaustive may be formed from ranges of quantities such as age, weight or income. The scales may involve time or frequency; they may involve measurements of money or counts.

## Figure 6. Pie Chart, Quantity Groups



Sometimes the groups are just names but given the question and answers they are exclusive and exhaustive.





Some categories are not naturally exclusive (e.g., personal values) so a single choice must be forced as indicated by a superlative (e.g., favorite, top, toughest).

# Figure 8. Pie Chart with Superlatives



When the slices of a pie are not exclusive or exhaustive and a forced choice is not indicated, we may conclude answers were limited to just one of those shown.

## Figure 9. Pie Chart with implied forced choices



When the slices in a percentage pie chart do not add to 100%, a plausible explanation is rounding error.

## Figure 10. Pie Charts Lacking a 100% Total



5. PERCENTAGE BAR GRAPHS

Bars in a bar graph can be any combination of exclusive and exhaustive and sum to any positive value.

If the bars in percentage bar graph are parts, then the sum of the bars may be restricted to a certain range depending on whether or not the bars are exclusive or exhaustive. Table 2 illustrates the ranges of these sums.

Table 2	. Range of	f Bar	Grap	ph S	Sums	if	Bars	are	Parts
---------	------------	-------	------	------	------	----	------	-----	-------

Exclusive	Exhaustive	Range
Yes	Yes	= 100%
Yes	No	< 100%
No	Yes	> 100%
No	No	Any

Given this relationship, classifying bar charts by the sum of their bars may give us some information on the nature of the bars (e.g., part, exclusive and exhaustive).

Bars in the following graphs were identified as parts or wholes based mainly on the grammar of the caption.

Of the 116 captions, 56 used percent or percentage grammar ("percent of," "half of", "majority of"), 19 used "part-in-whole" ratio grammar (six in ten), 18 used superlative grammar ("most" or "almost all"), 19

<sup>&</sup>lt;sup>5</sup> "Don't know" may be valuable in politics and business.

used rank grammar ("first choice", "top choice," "No.1"), two used subject-verb grammar that is similar to whole and part, and two just presented the subject.

Table 3 classifies the 116 percentage bar charts from *USA Today* online by their sum and part-whole status.

usie et but Gruphs sy sum und Fute (Thore Status						
<b>Percentage Bar Graphs</b>	Total	Parts	Wholes			
By Sum Total of Bars	116	97	19			
< 95%	41	34	7			
> 105%	44	34	10			
= 100%	14	14				
95% to 105% (not 100)	17	15	2			

Table 3. Bar Graphs by Sum and Part-Whole Status

Consider first the graphs where the bars are parts.

## 6. PART BARS: SUM = 100%

Consider first the bar charts where the bars are parts and add to exactly 100%. Typically the bars are exclusive and exhaustive. Being exclusive or exhaustive may be natural as illustrated in Figure 11.

## Figure 11. Bar Charts, Natural Groups



Knowing what choices are relevant depends on the whole – the subjects being classified. In the dog graph, the whole must be dog owners since zero is absent.

When the bars are not exclusive or exhaustive and the words indicating a restricted choice (e.g., 'favorite' or 'best') are absent, the 100% sum is evidence that the instructions made them exclusive and exhaustive.





So why use a 100% bar chart if the data could have been presented as a pie chart?

PRO: Bar charts allow longer captions and more visual variety for the groups. Pie charts get boring. CON: But it takes time to determine whether the bars add to 100%, whether the bars are parts (slices of a pie), and whether the bars are exclusive and exhaustive. Pie charts communicate all four (100%, part, exclusive and exhaustive) necessarily and immediately.

As will be seen, percentage bar charts lead to difficulty if not ambiguity. These 100% percentage bar graphs are just the first step.

#### 7. PART BARS: SUM ~100%

Consider percentage bar graphs where the bars are parts and sum to within 5% of 100% but exclude 100%. When the bars appear exclusive and exhaustive, one plausible reason for a small deviation (less than half a percent per bar) from 100% is rounding.

### Figure 13. Bars That Sum to 99%



If the bars appear exclusive and exhaustive and their sum deviates downward from 100% by more than rounding, one explanation is that the bars are nonexhaustive in another dimension (e.g., "no response.")

#### Figure 14. Bars That Sum Between 95% and 99%



8. PART BARS: SUM < 95%

Consider percentage bar graphs where the bars are parts and sum to less than 95%. Typically these bars are exclusive but are non-exhaustive.

## Figure 15. Bar Charts, Sum < 95%, Non-Exhaustive



A superlative (favorite or most) may indicate exclusivity but may not mean 'many' much less a 'majority.'

## Figure 16. Bars, Sum < 95%, Singular Superlative



A superlative (favorite or most) modifying a plural may have a different meaning. See discussion later.





If the bars are naturally exhaustive the missing category may be of a different kind or due to an error.

## Figure 18. Bar Chart, Sum < 95%, Exhaustive?





Consider percentage bar graphs where the bar percentages are parts and add to more than 105% (which usually eliminates rounding). Typically, categories are non-exclusive and non-exhaustive.

#### Figure 19. Bar Chart, Sum > 105%, Non-Exclusive



By Anne R. Carey and Gia Kereselidze, USA TODAY Source: Caloric Countries Why is non-exclusivity so common in these graphs?

PRO: Non-exclusive data is common when describing human values (where multiple choices are possible) and these graphs are designed to be of general interest. E.g., most foods we eat are not our 'favorite' food.

CON: Allowing multiple choices generates larger numbers. When describing a single bar percentage it matters whether the answers were exclusive or not. In addition, the rank of a bar may change depending on whether the answers are non-exclusive or exclusive.

Consider the graph in Figure 19. If the same respondent were asked for the *biggest* reason, the percentage who say "don't exercise" might drop from 59% to 9% and drop from second rank to last.

## 10. PART BARS: SUPERLATIVES & > 100%

When the bars sum to more than 100% and the categories are not exclusive, including a superlative in a title may seem like a contradiction since superlatives can indicate exclusivity.

## Figure 20. Bar Chart, Sum >105% with Superlative



But as shown in Figure 20, a bar graph may use "top" in the title yet have bars that are non-exclusive. The superlative refers to the most common answers given by all the respondents collectively rather than the one answer given by an individual respondent.

If the title includes a superlative (e.g., most, favorite, top) that modifies a singular (e.g., favorite food) and this reflects the reality of the question, then the answers should be exclusive. If the title includes a superlative that modifies a group or plural (e.g., favorite foods) then we can't tell if this refers to the question asked about one's favorite food, if this refers to the favorite foods of the respondents taken collectively, or both.

## Figure 21. Superlative in the Title



## 11. NON-EXCLUSIVE AND NON-EXHAUSTIVE

Bars that are parts can have any positive sum if they are non-exclusive and non-exhaustive. See Table 2.





It is inappropriate to add non-exclusive percentages. This is obvious when the total exceeds 100%. Adding non-exclusive percentages is not as obvious but is still inappropriate when the total is less than 100%. E.g., "30% of cars have a bumper sticker."

## **12. WHOLE BAR GRAPHS**

Now consider percentage bar charts where the bars are wholes. Although less common than charts where bars are wholes, these common-part bar charts permit comparisons that control for different sized groups.

How can we know if percentage bars are parts or wholes without reading the caption? Table 2 presented three conditions where combinations of factors (part, exclusivity, exhaustiveness) were sufficient to determine the range for the sum of the bar percentages.

In turn, these sums are necessary for these three conditions to exist when the bars are parts.

If a condition exists with a different sum than is necessary, then this is sufficient to conclude that some factor in the condition is different than assumed.

For example, suppose we assume the bars are the only parts but the combination of exclusive, exhaustive and sum factors fails in Table 2. If we are certain about exclusivity and exhaustiveness, then our assumption about the bars being the only parts must have been wrong and we may infer the bars are wholes.

If the sum is known, the violation of these three necessary conditions implies three sufficient conditions.

Table 4. Bar Graphs: D	<b>Determining Part or</b>	Whole
------------------------	----------------------------	-------

SUM*	Exclusive	Exhaustive	Part/Whole
≠ 100%	Yes	Yes	Whole
≥ 100%	Yes	No	Whole
≤ 100%	No	Yes	Whole

\* Allow up to 0.5% per bar for rounding error.

These sufficient conditions yield two rules involving simpler sufficient conditions which avoid problems with rounding.

#### Bars must be wholes if

- Rule #1: Sum > 105% and bars are exclusive or
- Rule #2: Sum < 95% and bars are exhaustive.

## 13. EXCLUSIVE WHOLE BARS: SUM > 105%

Consider bars that are naturally exclusive and sum to more than 105%. Regardless of whether the bars are exhaustive (e.g., sex) or not (e.g., places, dates or ages), these bars must be wholes.

#### Figure 23. Exclusive Whole Bars: Sum > 105%



#### 14. EXHAUSTIVE WHOLE BARS: SUM < 95%

Consider bars that are naturally exhaustive and sum to less than 95%. Regardless of whether they are exclusive (e.g., sex or marital status) or not (no example of such in these graphs), these bars must be wholes.

#### Figure 24. Exhaustive Whole Bars: Sum < 95%



### **15. IDENTIFYING PART AND WHOLE**

To be statistically literate in reading graphs, one cannot accept captions at face value. Mistakes are possible. Note the confusing caption of this graph. The first line excludes those who don't do spring cleaning, whereas the second line includes those who don't. Both statements may be true, but the second is irrelevant to this graph if the first statement is true.

## Figure 25. Graph with a Confusing Caption



00

To be statistically literate, one must look for consistency between the caption, the results of the aforementioned rules and results obtained from knowledge of the content. This requires identifying exclusivity and exhaustiveness, and identifying part and whole in ratios.

Table 5 compares the Rule Wholes predicted by the two rules with the Caption Wholes inferred from the grammatical roles of the terms in the captions. Nonrule categories (Other) are included for completeness.

Rule #1 worked perfectly: it predicted 10 and these were the same 10 obtained from the captions. Rule #2 did not work as well. It predicted 7 bar graphs as wholes, but three were actually parts and it failed to predict three whole-bar graphs within in its range. And there was one whole-bar graph very close to 100%.

Sum Range	Rule	Rule Wholes	Caption Wholes	Caption Parts
	ALL		18	98
> 100%	#1	10	10	
	Other		0	34
< 100%	#2	7	4	3
	Other		3	31
~ 100%	Other		1	16
= 100%	Other		0	14

 Table 5. Bar Graphs Sufficient for Whole Bars

The three graphs that Rule #2 predicted wrong (shaded in red) are in Figure 18 and Figure 27 (bottom). Each time, bars that seemed exhaustive were not.

The three graphs that Rule #2 missed (shaded in yellow) are in Figure 26 (both) and in Figure 30 (top). The whole-bar graph missed because it was too close to 100% in Figure 31 (bottom). All three of these graphs will be analyzed later in this paper.

#### **16. ANALYSIS**

Captions typically provided information on the partwhole status of percentage bars, but to be statistically literate we want ways to confirm their accuracy.

Graph titles were not helpful in indicating part-whole status. Titles focused on the topic or question: "Tried web dating." Ratio keywords (e.g., percentage, rate or chance) were not used. But 'among' appeared in one title while relative pronouns appeared in seven.<sup>6</sup> Each time the item indicated had the same part-whole status in the graph. Some subtitles indicated part-whole status (e.g., "Percent of shoplifters who are adults").

The two rules provide independent predictions, but their accuracy depends on correctly assessing exclusivity and exhaustiveness which in turn depends on the context. For example, these age groups (over 65 and 18-65) are exhaustive among adults but college majors are not exclusive if double-majors are allowed.

Unless other rules are found, informal guidelines involving the content may be required to confirm the part-whole status for most bar graphs. Content includes the nature of the things being counted as presented in the bar names, title, subtitle and caption.

A simple content-based rule might be 'impossibility.' If treating the bar one way (as part or whole) makes it impossible to create a percentage that is meaningful or relevant, the bar must be the other (whole or part).

## 17. WHY SOME BARS ARE WHOLES

Rule #2 failed to predict these bars as wholes: the groups are not exhaustive and sum to less than 100%.

## Figure 26. Non-Exhaustive Whole Bars: < 100%



Typically such bars are parts. But in both these graphs the bars are wholes. It doesn't make sense to say that, "28% of kids in a population are in 1980" or "20% of gym members live in Delaware."<sup>7</sup>

We could create a new rule: dates and places are always wholes (unless the percentages are exhaustive and add to 100%). But exceptions involve distribution (Figure 25) and rounding (bottom of Figure 30),

## **18. AMBIGUOUS THREE-FACTOR GRAPHS**

Even if the bars are identified as part or whole, the presence of three factors may lead to ambiguity.

#### Figure 27. Ambiguous Three-Factor Graphs



In the top graph, assuming the bars are non-exclusive parts, the whole can be either "men" or "men wanting a spa treatment."<sup>8</sup> Since these percentages seem too high for all men and the latter may give a higher percentage,

<sup>7</sup> If these three states are parts, then 40% of all gym members live in the remainder: e.g., California, New York and Texas.

<sup>&</sup>lt;sup>6</sup> AMONG: "Fighting among teens." WHO: "Guests who bring gifts", "People who chat and drive," "Students who speak up," Girls who get fit," "People who eat healthily" and "Americans who fish." THAT: "Things that go with coffee."

<sup>&</sup>lt;sup>8</sup> Percentage of "*men* who want a spa treatment massage" vs. "*spa-treatment wanting men* who want a massage."

 $P(AB|C) \le P(A|BC)^9$ , the latter is much more likely to be the whole than the former.

In the bottom graph, the bars are exclusive and add to less than 100%, so if the bars are the only parts they must not be exhaustive. But these areas of the US are exhaustive so Rule #2 predicted the bars as wholes. But notice the caption for this graph as shown in Figure 28.

Figure 28. Ambiguous Three-Factor Graph Caption

Nearly half (48%) of adults who drive and have a cellphone say they talk at least sometimes while driving; they are more likely to live in the



If this caption is correct, both "chat and drive" and geographical region are parts. One wonders if the caption is incorrect and the bars are indeed wholes.

As another example, consider the top graph in Figure 29 where the population is 'car occupants in major accidents.' Since the bars are exclusive and sum to more than 100%, the bars must be wholes. But which of the two factors in the title are parts? We can say, 50% of those were not buckled up and killed, 50% of those not buckled were killed or 50% of those killed were not buckled up. The latter is consistent with the caption: "... a higher percentage of people killed in pickup trucks crashes in 2003 were not buckled up ..."





In the bottom graph, we can't tell if "women" and "men" identify the sex of the business executives (part) or the sex of those surveyed (whole).

## **19. AMBIGUOUS PART-WHOLE BARS**

Even when there is no third factor, some graphs are ambiguous as to whether the bar is part or whole.

In the top graph in Figure 30, 'DUI' is an acronym for 'Driving Under the Influence' (typically alcohol). The age bars are exclusive and sum to 92%. These bars could be parts since the age groups may not be exhaustive (the age group of 35 and up is omitted). If so, then 29% of DUIs last year are for 21-25 year olds. But the bars could be wholes. If so, then 29% of 21-25 year olds had a DUI in the past year. The latter is consistent with the caption: *"Twenty-somethings are more likely than other adults to say they drove while intoxicated."* 





In the bottom graph, the 'continent' bars are exclusive, exhaustive and sum to 99%. The bars could be parts with one percent due to rounding or they could be wholes (unless the 56% is unrealistic ... ).

A "perfect storm" bar graph is one where the bars sum to almost 100% and where they can form meaningful statements being treated as either parts or wholes.

## Figure 31. Bar Chart, A Nearly "Perfect Storm"



If we allow 2% for rounding, we can't tell whether "15% of guests who bring gifts have low incomes" or "15% of guests with low income bring gifts." The

<sup>&</sup>lt;sup>9</sup> P(AB|C) = P(ABC)/P(C) and P(A|BC) = P(ABC)/P(BC).  $P(AB|C)/P(A|BC) = P(BC)/P(C) \le 1$  so  $P(AB|C) \le P(A|BC)$ .

latter agrees with the caption: "people with higher incomes are more apt to bring gifts."

## **20. CAPTIONS**

Captions all but determine a reader's ability to read and interpret percentage bar graphs. But captions can be ambiguous or misleading.

Does 'apt' mean the same as 'likely'? "People with higher incomes are more apt to bring gifts."

Does 'most likely' apply to individual choices or group choices? "More than half of adults think hybrid engines will most likely power vehicles in 2010."

## **21. STUDENT RESULTS**

Students who were trained on part-whole distinctions in tables were asked to read some of these graphs.

Half of 16 students thought the bars in the key graph on the top of Figure 30 were wholes and half thought they were parts. This indicates this graph is very hard to interpret without reading the caption.

Seven of 16 thought the bars in the sunflower graph in Figure 24 were parts or could not tell. Since many bars are parts, this may indicate students' bias: they just don't expect to see bars that are wholes.

Four of 15 thought the bars in the spa treatment graph in top of Figure 27 were exclusive. This may indicate student misunderstanding of 'exclusive.'

#### 22. RECOMMENDATIONS FOR EDITORS

Those who construct bar graphs of percentages should minimize ambiguity about the population surveyed, the issue and the results. Here are recommendations for copy editors who create titles and captions for percentage bar charts.

- 1. Develop a solid knowledge of the grammatical indicators for parts and wholes in percentages. Review descriptions and comparisons of multi-factor parts and wholes using percent, percentage, likely, ratio, superlative and rank grammars.
- 2. Ensure that the bar part-whole status is accurately described in the caption, the title or the subtitle.
- 3. Indicate when the bars are parts. Consider adding "distributed' in the title or subtitle when the bars are parts that should add to 100%.
- 4. Indicate when the bars are wholes. Consider adding "Average" in a subtitle or as a bar when appropriate.
- 5. Adjust pie charts (Figure 10) and 100% bar charts (Figure 13 and Figure 14) for rounding so the pieces add to 100%. To minimize the impact of the adjustment, apply the difference to either the largest piece(s) or to the piece(s) closest to the half-way point.
- 6. Identify when answers are non-exclusive. Consider adding "Multiple answers permitted" in a subtitle.
- 7. Indicate when answers are non-exhaustive. Consider putting "Some answers not shown" in a subti-

tle or adding "Other" as a category to make partbars exhaustive.

8. Don't rely on superlatives (most, best) to indicate a forced choice for each individual. Superlatives often reflect the most common responses for a group.

#### 23. CONCLUSION

Reading and interpreting data presented in tables and graphs is an essential aspect of statistical literacy.

Graphs show relationships; tables present data. Graphs seem simpler than tables because the relationships are visible. But having visible relationships doesn't mean the items involved (the population, exclusivity and part vs. whole) are easier to understand.

To better assess statistical literacy as suggested by the GAISE college guidelines, educators should study student difficulties in reading percentages in graphs.

This kind of analysis should be extended to other publications that may be written for a more sophisticated audience (e.g., Time, Newsweek, the Wall Street Journal and The Economist) where there may be a different mix of graphs: more percentage change/difference comparisons, more line graphs and more scatter-plots.

#### REFERENCES

ASA Guidelines for Assessment and Instruction in Statistics Education (GAISE, 2006).<sup>10</sup>

Schield, M. (2000). *Statistical Literacy: Reading Ratios of Rates and Percentages.*<sup>11</sup> 2000 ASA Proceedings of the Section on Statistical Education, p. 76-81.

Schield, M. (2001).<sup>12</sup> Statistical Literacy: Reading Tables of Rates and Percentages. 2001 ASA Proceedings of the Section on Statistical Education.

Schield, M. (2004).<sup>13</sup> Statistical Literacy and Liberal Education at Augsburg College. Peer Review Summer.

Schield, M. (2006).<sup>14</sup> Statistical Literacy Survey Analysis: Reading Graphs and Tables of Rates and Percentages. ICOTS conference.

Acknowledgments: This research was conducted under the W. M. Keck Statistical Literacy Project dedicated "to support the development of statistical literacy as an interdisciplinary discipline in the liberal arts." Thanks to Donald Macnaughton, Tom Burnham, Marc Isaacson and Cynthia Schield for their suggestions.

**Contact:** Schield is at Schield@Augsburg.edu. This paper is at <u>www.StatLit.org/pdf/2006SchieldASA.pdf</u>.

<sup>&</sup>lt;sup>10</sup> Copy at <u>www.amstat.org/education/gaise/</u>

<sup>&</sup>lt;sup>11</sup> Copy at <u>www.StatLit.org/pdf/2000SchieldASA.pdf</u>.

<sup>&</sup>lt;sup>12</sup> Copy at <u>www.StatLit.org/pdf/2001SchieldASA.pdf</u>.

<sup>&</sup>lt;sup>13</sup> Copy at www.StatLit.org/pdf/2004SchieldAACU.pdf.

<sup>&</sup>lt;sup>14</sup> Copy at <u>www.StatLit.org/pdf/2006SchieldICOTS.pdf</u>.