Rethinking (and challenging) the Ranks of Visual Channels

Reading Group • Friday, March 18, 2022

Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods

WILLIAM S. CLEVELAND and ROBERT McGILL*

The subject of graphical methods for data analysis and for data presentation needs a scientific foundation. In this article we take a few steps in the direction of establishing such a foundation. Our approach is based on *graphical perception*—the visual decoding of information encoded on graphs—and it includes both theory and experimentation to test the theory. The theory deals with a small but important piece of the whole process of graphical perception. The first part is an identification of a set of *elementary perceptual tasks* that are carried out when people extract quantitative information from graphs. The second part is an ordering of the tasks on the basis of how accurately people perform them. Elements of the theory are tested by experimentation in which subjects record their judgments of the quantitative information on graphs. The experiments validate these elements but also suggest that the set of elementary tasks should be expanded. The theory provides a guideline for graph construction: Graphs should employ elementary tasks as high in the ordering as possible. This principle is applied to a variety of graphs, including bar charts, divided bar charts, pie charts, and statistical maps with shading. The conclusion is that radical surgery on these popular graphs is needed, and as replacements we offer alternative graphical forms—*dot charts, dot charts with grouping*, and *framed-rectangle charts*.

KEY WORDS: Computer graphics; Psychophysics.

1. INTRODUCTION

Nearly 200 years ago William Playfair (1786) began the serious use of graphs for looking at data. More than 50 years ago a battle raged on the pages of the *Journal of the American Statistical Association* about the relative merits of bar charts and pie charts (Eells 1926; Croxton 1927; Croxton and Stryker 1927; von Huhn 1927). Today graphs are a vital part of statistical data analysis and a vital part of communication in science and technology, business, education, and the mass media.

Still, graph design for data analysis and presentation is largely unscientific. This is why Cox (1978) argued, "There is a major need for a theory of graphical methods" (p. 5), and why Kruskal (1975) stated "in choosing, constructing, and comparing graphical methods we have little to go on but intuition, rule of thumb, and a kind of master-to-apprentice passing along of information. . . . there is neither theory nor systematic body of experiment as a guide" (p. 28–29).

There is, of course, much good common sense about how to make a graph. There are many treatises on graph construction (e.g., Schmid and Schmid 1979), bad practice has been uncovered (e.g., Tufte 1983), graphic designers certainly have shown us how to make a graph appealing to the eye (e.g., Marcus et al. 1980), statisticians have thought intensely about graphical methods for data analysis (e.g., Tukey 1977; Chambers et al. 1983), and cartographers have devoted great energy to the construction of statistical maps (Bertin 1973; Robinson, Sale, and Morrison 1978). The ANSI manual on time series charts (American National Standards Institute 1979) provides guidelines for making graphs, but the manual admits, "This standard . . . sets forth the best current usage, and offers standards 'by general agreement' rather than 'by scientific test' " (p. iii).

In this article we approach the science of graphs through human graphical perception. Our approach includes both theory and experimentation to test it.

The first part of the theory is a list of elementary perceptual tasks that people perform in extracting quantitative information from graphs. In the second part we hypothesize an ordering of the elementary tasks based on how accurately people perform them. We do not argue that this accuracy of quantitative extraction is the only aspect of a graph for which one might want to develop a theory, but it is an important one.

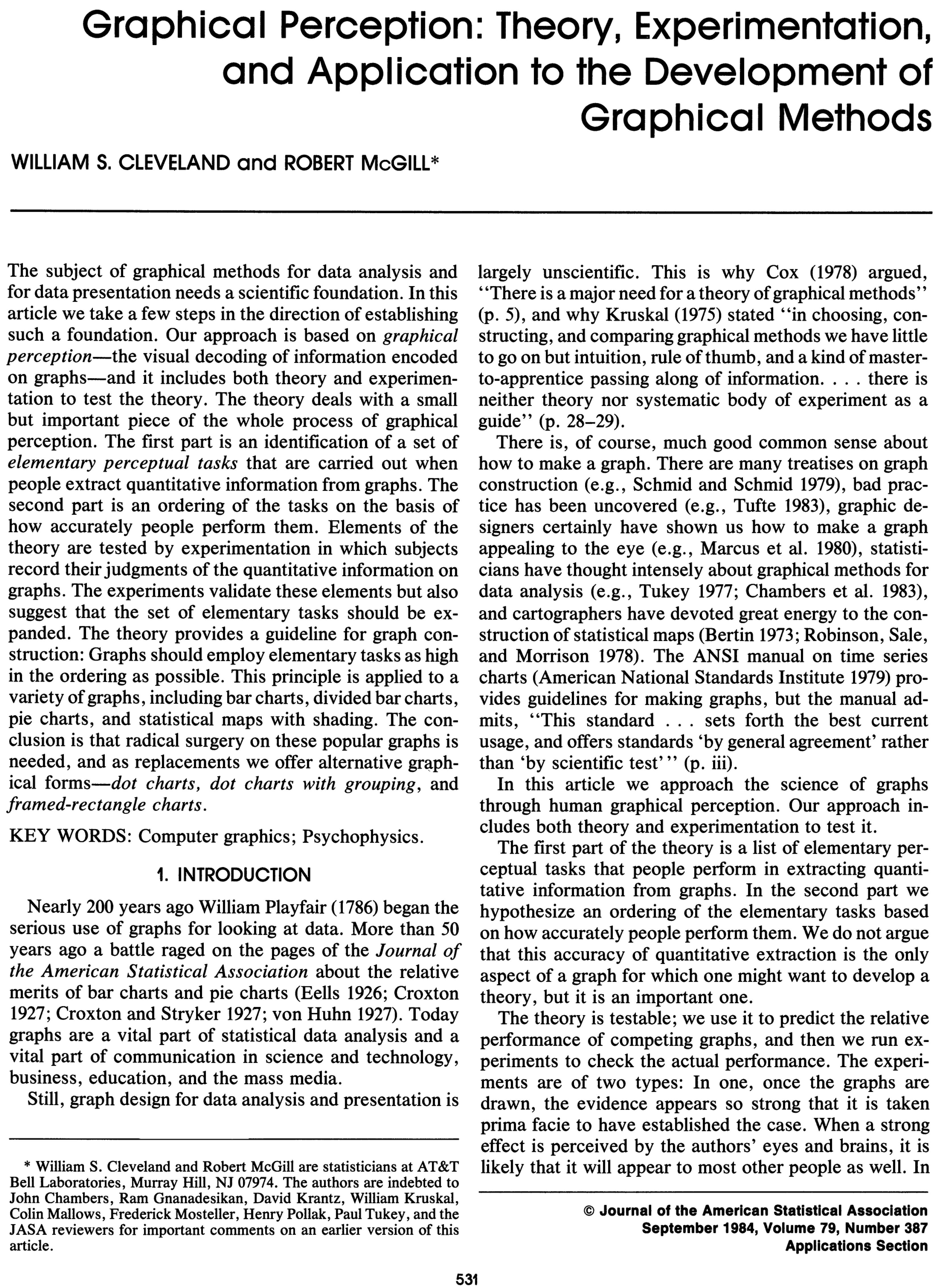
The theory is testable; we use it to predict the relative performance of competing graphs, and then we run experiments to check the actual performance. The experiments are of two types: In one, once the graphs are drawn, the evidence appears so strong that it is taken prima facie to have established the case. When a strong effect is perceived by the authors' eyes and brains, it is likely that it will appear to most other people as well. In

* William S. Cleveland and Robert McGill are statisticians at AT&T Bell Laboratories, Murray Hill, NJ 07974. The authors are indebted to John Chambers, Ram Gnanadesikan, David Krantz, William Kruskal, Colin Mallows, Frederick Mosteller, Henry Pollak, Paul Tukey, and the JASA reviewers for important comments on an earlier version of this article.

© Journal of the American Statistical Association
September 1984, Volume 79, Number 387
Applications Section

Cleveland & McGill's Ranking (1984)

↑ "Better"



Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods

WILLIAM S. CLEVELAND and ROBERT McGILL*

The subject of graphical methods for data analysis and for data presentation needs a scientific foundation. In this article we take a few steps in the direction of establishing such a foundation. Our approach is based on *graphical perception*—the visual decoding of information encoded on graphs—and it includes both theory and experimentation to test the theory. The theory deals with a small but important piece of the whole process of graphical perception. The first part is an identification of a set of *elementary perceptual tasks* that are carried out when people extract quantitative information from graphs. The second part is an ordering of the tasks on the basis of how accurately people perform them. Elements of the theory are tested by experimentation in which subjects record their judgments of the quantitative information on graphs. The experiments validate these elements but also suggest that the set of elementary tasks should be expanded. The theory provides a guideline for graph construction: Graphs should employ elementary tasks as high in the ordering as possible. This principle is applied to a variety of graphs, including bar charts, divided bar charts, pie charts, and statistical maps with shading. The conclusion is that radical surgery on these popular graphs is needed, and as replacements we offer alternative graphical forms—*dot charts, dot charts with grouping*, and *framed-rectangle charts*.

KEY WORDS: Computer graphics; Psychophysics.

1. INTRODUCTION

Nearly 200 years ago William Playfair (1786) began the serious use of graphs for looking at data. More than 50 years ago a battle raged on the pages of the *Journal of the American Statistical Association* about the relative merits of bar charts and pie charts (Eells 1926; Croxton 1927; Croxton and Stryker 1927; von Huhn 1927). Today graphs are a vital part of statistical data analysis and a vital part of communication in science and technology, business, education, and the mass media.

Still, graph design for data analysis and presentation is largely unscientific. This is why Cox (1978) argued, "There is a major need for a theory of graphical methods" (p. 5), and why Kruskal (1975) stated "in choosing, constructing, and comparing graphical methods we have little to go on but intuition, rule of thumb, and a kind of master-to-apprentice passing along of information. . . . there is neither theory nor systematic body of experiment as a guide" (p. 28–29).

There is, of course, much good common sense about how to make a graph. There are many treatises on graph construction (e.g., Schmid and Schmid 1979), bad practice has been uncovered (e.g., Tufte 1983), graphic designers certainly have shown us how to make a graph appealing to the eye (e.g., Marcus et al. 1980), statisticians have thought intensely about graphical methods for data analysis (e.g., Tukey 1977; Chambers et al. 1983), and cartographers have devoted great energy to the construction of statistical maps (Bertin 1973; Robinson, Sale, and Morrison 1978). The ANSI manual on time series charts (American National Standards Institute 1979) provides guidelines for making graphs, but the manual admits, "This standard . . . sets forth the best current usage, and offers standards 'by general agreement' rather than 'by scientific test' " (p. iii).

In this article we approach the science of graphs through human graphical perception. Our approach includes both theory and experimentation to test it.

The first part of the theory is a list of elementary perceptual tasks that people perform in extracting quantitative information from graphs. In the second part we hypothesize an ordering of the elementary tasks based on how accurately people perform them. We do not argue that this accuracy of quantitative extraction is the only aspect of a graph for which one might want to develop a theory, but it is an important one.

The theory is testable; we use it to predict the relative performance of competing graphs, and then we run experiments to check the actual performance. The experiments are of two types: In one, once the graphs are drawn, the evidence appears so strong that it is taken prima facie to have established the case. When a strong effect is perceived by the authors' eyes and brains, it is likely that it will appear to most other people as well. In

* William S. Cleveland and Robert McGill are statisticians at AT&T Bell Laboratories, Murray Hill, NJ 07974. The authors are indebted to John Chambers, Ram Gnanadesikan, David Krantz, William Kruskal, Colin Mallows, Frederick Mosteller, Henry Pollak, Paul Tukey, and the JASA reviewers for important comments on an earlier version of this article.

© Journal of the American Statistical Association
September 1984, Volume 79, Number 387
Applications Section

Cleveland & McGill's Ranking (1984)

↑ "Better"

1. Position (common scale) → Scatterplot

Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods

WILLIAM S. CLEVELAND and ROBERT McGILL*

The subject of graphical methods for data analysis and for data presentation needs a scientific foundation. In this article we take a few steps in the direction of establishing such a foundation. Our approach is based on *graphical perception*—the visual decoding of information encoded on graphs—and it includes both theory and experimentation to test the theory. The theory deals with a small but important piece of the whole process of graphical perception. The first part is an identification of a set of *elementary perceptual tasks* that are carried out when people extract quantitative information from graphs. The second part is an ordering of the tasks on the basis of how accurately people perform them. Elements of the theory are tested by experimentation in which subjects record their judgments of the quantitative information on graphs. The experiments validate these elements but also suggest that the set of elementary tasks should be expanded. The theory provides a guideline for graph construction: Graphs should employ elementary tasks as high in the ordering as possible. This principle is applied to a variety of graphs, including bar charts, divided bar charts, pie charts, and statistical maps with shading. The conclusion is that radical surgery on these popular graphs is needed, and as replacements we offer alternative graphical forms—*dot charts, dot charts with grouping*, and *framed-rectangle charts*.

KEY WORDS: Computer graphics; Psychophysics.

1. INTRODUCTION

Nearly 200 years ago William Playfair (1786) began the serious use of graphs for looking at data. More than 50 years ago a battle raged on the pages of the *Journal of the American Statistical Association* about the relative merits of bar charts and pie charts (Eells 1926; Croxton 1927; Croxton and Stryker 1927; von Huhn 1927). Today graphs are a vital part of statistical data analysis and a vital part of communication in science and technology, business, education, and the mass media.

Still, graph design for data analysis and presentation is largely unscientific. This is why Cox (1978) argued, "There is a major need for a theory of graphical methods" (p. 5), and why Kruskal (1975) stated "in choosing, constructing, and comparing graphical methods we have little to go on but intuition, rule of thumb, and a kind of master-to-apprentice passing along of information. . . . there is neither theory nor systematic body of experiment as a guide" (p. 28–29).

There is, of course, much good common sense about how to make a graph. There are many treatises on graph construction (e.g., Schmid and Schmid 1979), bad practice has been uncovered (e.g., Tufte 1983), graphic designers certainly have shown us how to make a graph appealing to the eye (e.g., Marcus et al. 1980), statisticians have thought intensely about graphical methods for data analysis (e.g., Tukey 1977; Chambers et al. 1983), and cartographers have devoted great energy to the construction of statistical maps (Bertin 1973; Robinson, Sale, and Morrison 1978). The ANSI manual on time series charts (American National Standards Institute 1979) provides guidelines for making graphs, but the manual admits, "This standard . . . sets forth the best current usage, and offers standards 'by general agreement' rather than 'by scientific test' " (p. iii).

In this article we approach the science of graphs through human graphical perception. Our approach includes both theory and experimentation to test it.

The first part of the theory is a list of elementary perceptual tasks that people perform in extracting quantitative information from graphs. In the second part we hypothesize an ordering of the elementary tasks based on how accurately people perform them. We do not argue that this accuracy of quantitative extraction is the only aspect of a graph for which one might want to develop a theory, but it is an important one.

The theory is testable; we use it to predict the relative performance of competing graphs, and then we run experiments to check the actual performance. The experiments are of two types: In one, once the graphs are drawn, the evidence appears so strong that it is taken prima facie to have established the case. When a strong effect is perceived by the authors' eyes and brains, it is likely that it will appear to most other people as well. In

* William S. Cleveland and Robert McGill are statisticians at AT&T Bell Laboratories, Murray Hill, NJ 07974. The authors are indebted to John Chambers, Ram Gnanadesikan, David Krantz, William Kruskal, Colin Mallows, Frederick Mosteller, Henry Pollak, Paul Tukey, and the JASA reviewers for important comments on an earlier version of this article.

© Journal of the American Statistical Association
September 1984, Volume 79, Number 387
Applications Section

Cleveland & McGill's Ranking (1984)

↑ "Better"

1. Position (common scale) → Scatterplot
2. Position (non-aligned scales) → Multiple scatterplots

Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods

WILLIAM S. CLEVELAND and ROBERT McGILL*

The subject of graphical methods for data analysis and for data presentation needs a scientific foundation. In this article we take a few steps in the direction of establishing such a foundation. Our approach is based on *graphical perception*—the visual decoding of information encoded on graphs—and it includes both theory and experimentation to test the theory. The theory deals with a small but important piece of the whole process of graphical perception. The first part is an identification of a set of *elementary perceptual tasks* that are carried out when people extract quantitative information from graphs. The second part is an ordering of the tasks on the basis of how accurately people perform them. Elements of the theory are tested by experimentation in which subjects record their judgments of the quantitative information on graphs. The experiments validate these elements but also suggest that the set of elementary tasks should be expanded. The theory provides a guideline for graph construction: Graphs should employ elementary tasks as high in the ordering as possible. This principle is applied to a variety of graphs, including bar charts, divided bar charts, pie charts, and statistical maps with shading. The conclusion is that radical surgery on these popular graphs is needed, and as replacements we offer alternative graphical forms—*dot charts, dot charts with grouping*, and *framed-rectangle charts*.

KEY WORDS: Computer graphics; Psychophysics.

1. INTRODUCTION

Nearly 200 years ago William Playfair (1786) began the serious use of graphs for looking at data. More than 50 years ago a battle raged on the pages of the *Journal of the American Statistical Association* about the relative merits of bar charts and pie charts (Eells 1926; Croxton 1927; Croxton and Stryker 1927; von Huhn 1927). Today graphs are a vital part of statistical data analysis and a vital part of communication in science and technology, business, education, and the mass media.

Still, graph design for data analysis and presentation is largely unscientific. This is why Cox (1978) argued, "There is a major need for a theory of graphical methods" (p. 5), and why Kruskal (1975) stated "in choosing, constructing, and comparing graphical methods we have little to go on but intuition, rule of thumb, and a kind of master-to-apprentice passing along of information. . . . there is neither theory nor systematic body of experiment as a guide" (p. 28–29).

There is, of course, much good common sense about how to make a graph. There are many treatises on graph construction (e.g., Schmid and Schmid 1979), bad practice has been uncovered (e.g., Tufte 1983), graphic designers certainly have shown us how to make a graph appealing to the eye (e.g., Marcus et al. 1980), statisticians have thought intensely about graphical methods for data analysis (e.g., Tukey 1977; Chambers et al. 1983), and cartographers have devoted great energy to the construction of statistical maps (Bertin 1973; Robinson, Sale, and Morrison 1978). The ANSI manual on time series charts (American National Standards Institute 1979) provides guidelines for making graphs, but the manual admits, "This standard . . . sets forth the best current usage, and offers standards 'by general agreement' rather than 'by scientific test' " (p. iii).

In this article we approach the science of graphs through human graphical perception. Our approach includes both theory and experimentation to test it.

The first part of the theory is a list of elementary perceptual tasks that people perform in extracting quantitative information from graphs. In the second part we hypothesize an ordering of the elementary tasks based on how accurately people perform them. We do not argue that this accuracy of quantitative extraction is the only aspect of a graph for which one might want to develop a theory, but it is an important one.

The theory is testable; we use it to predict the relative performance of competing graphs, and then we run experiments to check the actual performance. The experiments are of two types: In one, once the graphs are drawn, the evidence appears so strong that it is taken prima facie to have established the case. When a strong effect is perceived by the authors' eyes and brains, it is likely that it will appear to most other people as well. In

* William S. Cleveland and Robert McGill are statisticians at AT&T Bell Laboratories, Murray Hill, NJ 07974. The authors are indebted to John Chambers, Ram Gnanadesikan, David Krantz, William Kruskal, Colin Mallows, Frederick Mosteller, Henry Pollak, Paul Tukey, and the JASA reviewers for important comments on an earlier version of this article.

© Journal of the American Statistical Association
September 1984, Volume 79, Number 387
Applications Section

Cleveland & McGill's Ranking (1984)

↑ "Better"

1. Position (common scale) → Scatterplot
2. Position (non-aligned scales) → Multiple scatterplots
3. Length → Bar chart (🧂) or Gantt chart

Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods

WILLIAM S. CLEVELAND and ROBERT McGILL*

The subject of graphical methods for data analysis and for data presentation needs a scientific foundation. In this article we take a few steps in the direction of establishing such a foundation. Our approach is based on *graphical perception*—the visual decoding of information encoded on graphs—and it includes both theory and experimentation to test the theory. The theory deals with a small but important piece of the whole process of graphical perception. The first part is an identification of a set of *elementary perceptual tasks* that are carried out when people extract quantitative information from graphs. The second part is an ordering of the tasks on the basis of how accurately people perform them. Elements of the theory are tested by experimentation in which subjects record their judgments of the quantitative information on graphs. The experiments validate these elements but also suggest that the set of elementary tasks should be expanded. The theory provides a guideline for graph construction: Graphs should employ elementary tasks as high in the ordering as possible. This principle is applied to a variety of graphs, including bar charts, divided bar charts, pie charts, and statistical maps with shading. The conclusion is that radical surgery on these popular graphs is needed, and as replacements we offer alternative graphical forms—*dot charts, dot charts with grouping*, and *framed-rectangle charts*.

KEY WORDS: Computer graphics; Psychophysics.

1. INTRODUCTION

Nearly 200 years ago William Playfair (1786) began the serious use of graphs for looking at data. More than 50 years ago a battle raged on the pages of the *Journal of the American Statistical Association* about the relative merits of bar charts and pie charts (Eells 1926; Croxton 1927; Croxton and Stryker 1927; von Huhn 1927). Today graphs are a vital part of statistical data analysis and a vital part of communication in science and technology, business, education, and the mass media.

Still, graph design for data analysis and presentation is largely unscientific. This is why Cox (1978) argued, "There is a major need for a theory of graphical methods" (p. 5), and why Kruskal (1975) stated "in choosing, constructing, and comparing graphical methods we have little to go on but intuition, rule of thumb, and a kind of master-to-apprentice passing along of information. . . . there is neither theory nor systematic body of experiment as a guide" (p. 28–29).

There is, of course, much good common sense about how to make a graph. There are many treatises on graph construction (e.g., Schmid and Schmid 1979), bad practice has been uncovered (e.g., Tufte 1983), graphic designers certainly have shown us how to make a graph appealing to the eye (e.g., Marcus et al. 1980), statisticians have thought intensely about graphical methods for data analysis (e.g., Tukey 1977; Chambers et al. 1983), and cartographers have devoted great energy to the construction of statistical maps (Bertin 1973; Robinson, Sale, and Morrison 1978). The ANSI manual on time series charts (American National Standards Institute 1979) provides guidelines for making graphs, but the manual admits, "This standard . . . sets forth the best current usage, and offers standards 'by general agreement' rather than 'by scientific test'" (p. iii).

In this article we approach the science of graphs through human graphical perception. Our approach includes both theory and experimentation to test it.

The first part of the theory is a list of elementary perceptual tasks that people perform in extracting quantitative information from graphs. In the second part we hypothesize an ordering of the elementary tasks based on how accurately people perform them. We do not argue that this accuracy of quantitative extraction is the only aspect of a graph for which one might want to develop a theory, but it is an important one.

The theory is testable; we use it to predict the relative performance of competing graphs, and then we run experiments to check the actual performance. The experiments are of two types: In one, once the graphs are drawn, the evidence appears so strong that it is taken prima facie to have established the case. When a strong effect is perceived by the authors' eyes and brains, it is likely that it will appear to most other people as well. In

* William S. Cleveland and Robert McGill are statisticians at AT&T Bell Laboratories, Murray Hill, NJ 07974. The authors are indebted to John Chambers, Ram Gnanadesikan, David Krantz, William Kruskal, Colin Mallows, Frederick Mosteller, Henry Pollak, Paul Tukey, and the JASA reviewers for important comments on an earlier version of this article.

© Journal of the American Statistical Association
September 1984, Volume 79, Number 387
Applications Section

Cleveland & McGill's Ranking (1984)

↑ "Better"

1. Position (common scale) → Scatterplot
2. Position (non-aligned scales) → Multiple scatterplots
3. Length → Bar chart (🧂) or Gantt chart
4. Angle → Pie chart

Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods

WILLIAM S. CLEVELAND and ROBERT McGILL*

The subject of graphical methods for data analysis and for data presentation needs a scientific foundation. In this article we take a few steps in the direction of establishing such a foundation. Our approach is based on *graphical perception*—the visual decoding of information encoded on graphs—and it includes both theory and experimentation to test the theory. The theory deals with a small but important piece of the whole process of graphical perception. The first part is an identification of a set of *elementary perceptual tasks* that are carried out when people extract quantitative information from graphs. The second part is an ordering of the tasks on the basis of how accurately people perform them. Elements of the theory are tested by experimentation in which subjects record their judgments of the quantitative information on graphs. The experiments validate these elements but also suggest that the set of elementary tasks should be expanded. The theory provides a guideline for graph construction: Graphs should employ elementary tasks as high in the ordering as possible. This principle is applied to a variety of graphs, including bar charts, divided bar charts, pie charts, and statistical maps with shading. The conclusion is that radical surgery on these popular graphs is needed, and as replacements we offer alternative graphical forms—*dot charts, dot charts with grouping*, and *framed-rectangle charts*.

KEY WORDS: Computer graphics; Psychophysics.

1. INTRODUCTION

Nearly 200 years ago William Playfair (1786) began the serious use of graphs for looking at data. More than 50 years ago a battle raged on the pages of the *Journal of the American Statistical Association* about the relative merits of bar charts and pie charts (Eells 1926; Croxton 1927; Croxton and Stryker 1927; von Huhn 1927). Today graphs are a vital part of statistical data analysis and a vital part of communication in science and technology, business, education, and the mass media.

Still, graph design for data analysis and presentation is largely unscientific. This is why Cox (1978) argued, "There is a major need for a theory of graphical methods" (p. 5), and why Kruskal (1975) stated "in choosing, constructing, and comparing graphical methods we have little to go on but intuition, rule of thumb, and a kind of master-to-apprentice passing along of information. . . . there is neither theory nor systematic body of experiment as a guide" (p. 28–29).

There is, of course, much good common sense about how to make a graph. There are many treatises on graph construction (e.g., Schmid and Schmid 1979), bad practice has been uncovered (e.g., Tufte 1983), graphic designers certainly have shown us how to make a graph appealing to the eye (e.g., Marcus et al. 1980), statisticians have thought intensely about graphical methods for data analysis (e.g., Tukey 1977; Chambers et al. 1983), and cartographers have devoted great energy to the construction of statistical maps (Bertin 1973; Robinson, Sale, and Morrison 1978). The ANSI manual on time series charts (American National Standards Institute 1979) provides guidelines for making graphs, but the manual admits, "This standard . . . sets forth the best current usage, and offers standards 'by general agreement' rather than 'by scientific test' " (p. iii).

In this article we approach the science of graphs through human graphical perception. Our approach includes both theory and experimentation to test it.

The first part of the theory is a list of elementary perceptual tasks that people perform in extracting quantitative information from graphs. In the second part we hypothesize an ordering of the elementary tasks based on how accurately people perform them. We do not argue that this accuracy of quantitative extraction is the only aspect of a graph for which one might want to develop a theory, but it is an important one.

The theory is testable; we use it to predict the relative performance of competing graphs, and then we run experiments to check the actual performance. The experiments are of two types: In one, once the graphs are drawn, the evidence appears so strong that it is taken prima facie to have established the case. When a strong effect is perceived by the authors' eyes and brains, it is likely that it will appear to most other people as well. In

* William S. Cleveland and Robert McGill are statisticians at AT&T Bell Laboratories, Murray Hill, NJ 07974. The authors are indebted to John Chambers, Ram Gnanadesikan, David Krantz, William Kruskal, Colin Mallows, Frederick Mosteller, Henry Pollak, Paul Tukey, and the JASA reviewers for important comments on an earlier version of this article.

© Journal of the American Statistical Association
September 1984, Volume 79, Number 387
Applications Section

Cleveland & McGill's Ranking (1984)

↑ "Better"

1. Position (common scale) → Scatterplot
2. Position (non-aligned scales) → Multiple scatterplots
3. Length → Bar chart (🧂) or Gantt chart
4. Angle → Pie chart
5. Area → Bubble chart

Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods

WILLIAM S. CLEVELAND and ROBERT McGILL*

The subject of graphical methods for data analysis and for data presentation needs a scientific foundation. In this article we take a few steps in the direction of establishing such a foundation. Our approach is based on *graphical perception*—the visual decoding of information encoded on graphs—and it includes both theory and experimentation to test the theory. The theory deals with a small but important piece of the whole process of graphical perception. The first part is an identification of a set of *elementary perceptual tasks* that are carried out when people extract quantitative information from graphs. The second part is an ordering of the tasks on the basis of how accurately people perform them. Elements of the theory are tested by experimentation in which subjects record their judgments of the quantitative information on graphs. The experiments validate these elements but also suggest that the set of elementary tasks should be expanded. The theory provides a guideline for graph construction: Graphs should employ elementary tasks as high in the ordering as possible. This principle is applied to a variety of graphs, including bar charts, divided bar charts, pie charts, and statistical maps with shading. The conclusion is that radical surgery on these popular graphs is needed, and as replacements we offer alternative graphical forms—*dot charts, dot charts with grouping*, and *framed-rectangle charts*.

KEY WORDS: Computer graphics; Psychophysics.

1. INTRODUCTION

Nearly 200 years ago William Playfair (1786) began the serious use of graphs for looking at data. More than 50 years ago a battle raged on the pages of the *Journal of the American Statistical Association* about the relative merits of bar charts and pie charts (Eells 1926; Croxton 1927; Croxton and Stryker 1927; von Huhn 1927). Today graphs are a vital part of statistical data analysis and a vital part of communication in science and technology, business, education, and the mass media.

Still, graph design for data analysis and presentation is largely unscientific. This is why Cox (1978) argued, "There is a major need for a theory of graphical methods" (p. 5), and why Kruskal (1975) stated "in choosing, constructing, and comparing graphical methods we have little to go on but intuition, rule of thumb, and a kind of master-to-apprentice passing along of information. . . . there is neither theory nor systematic body of experiment as a guide" (p. 28–29).

There is, of course, much good common sense about how to make a graph. There are many treatises on graph construction (e.g., Schmid and Schmid 1979), bad practice has been uncovered (e.g., Tufte 1983), graphic designers certainly have shown us how to make a graph appealing to the eye (e.g., Marcus et al. 1980), statisticians have thought intensely about graphical methods for data analysis (e.g., Tukey 1977; Chambers et al. 1983), and cartographers have devoted great energy to the construction of statistical maps (Bertin 1973; Robinson, Sale, and Morrison 1978). The ANSI manual on time series charts (American National Standards Institute 1979) provides guidelines for making graphs, but the manual admits, "This standard . . . sets forth the best current usage, and offers standards 'by general agreement' rather than 'by scientific test'" (p. iii).

In this article we approach the science of graphs through human graphical perception. Our approach includes both theory and experimentation to test it.

The first part of the theory is a list of elementary perceptual tasks that people perform in extracting quantitative information from graphs. In the second part we hypothesize an ordering of the elementary tasks based on how accurately people perform them. We do not argue that this accuracy of quantitative extraction is the only aspect of a graph for which one might want to develop a theory, but it is an important one.

The theory is testable; we use it to predict the relative performance of competing graphs, and then we run experiments to check the actual performance. The experiments are of two types: In one, once the graphs are drawn, the evidence appears so strong that it is taken prima facie to have established the case. When a strong effect is perceived by the authors' eyes and brains, it is likely that it will appear to most other people as well. In

* William S. Cleveland and Robert McGill are statisticians at AT&T Bell Laboratories, Murray Hill, NJ 07974. The authors are indebted to John Chambers, Ram Gnanadesikan, David Krantz, William Kruskal, Colin Mallows, Frederick Mosteller, Henry Pollak, Paul Tukey, and the JASA reviewers for important comments on an earlier version of this article.

© Journal of the American Statistical Association
September 1984, Volume 79, Number 387
Applications Section

Cleveland & McGill's Ranking (1984)

↑ "Better"

1. Position (common scale) → Scatterplot
2. Position (non-aligned scales) → Multiple scatterplots
3. Length → Bar chart (🧂) or Gantt chart
4. Angle → Pie chart
5. Area → Bubble chart
6. Color luminance → Heatmap

Cleveland & McGill’s Ranking (1984)

DEFINITION

Ranking of the most perceptually precise visual channels estimated from two-value ratio judgment tasks for Data Visualization

"What percentage is the smaller value compared to the larger?"

Cleveland & McGill's Ranking (1984)

DEFINITION

Ranking of the most perceptually precise visual channels estimated from two-value ratio judgment tasks for Data Visualization

"What percentage is the smaller value compared to the larger?"

ASSUMPTION

This ranking should hold for different tasks and numbers of marks

Cleveland & McGill’s Ranking (1984)

DEFINITION

Ranking of the most perceptually precise visual channels estimated from two-value ratio judgment tasks for Data Visualization

"What percentage is the smaller value compared to the larger?"

ASSUMPTION

This ranking should hold for different tasks and numbers of marks

QUESTIONS

- Does this ranking hold for other contexts?
- Can we use this ranking as guidelines for Data Visualization?

Rankings

Cleveland & McGill’s Ranking

Position

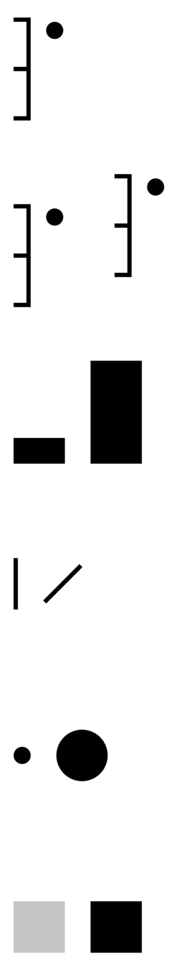
Position

Length

Angle

Area

Color luminance



Rankings

Cleveland & McGill's Ranking vs. Probabilistic Rankings

	BIAS				PRECISION				ERROR OF INDIVIDUAL RESPONSE			
Position |·	Angle	Length	Length	Length	Angle	Area	Area	Area	Angle	Bar	Bar	Bar
Position |· |·	Length	Bar	Bar	Bar	Area	Angle	Bar	Bar	Bar	Angle	Length	Length
Length	Bar	Angle	Lumin.	Lumin.	Lumin.	Lumin.	Angle	Angle	Length	Length	Lumin.	Lumin.
Angle	Lumin.	Lumin.	Angle	Line	Length	Bar	Lumin.	Lumin.	Lumin.	Lumin.	Angle	Area
Area	Area	Area	Line	Area	Bar	Length	Length	Length	Area	Area	Area	Angle
Color luminance	Line	Line	Area	Angle	Line	Line	Line	Line	Line	Line	Line	Line
	2	4	6	8	2	4	6	8	2	4	6	8

Rethinking the Ranks of Visual Channels

PAPER

- Caitlyn M. McColeman, Fumeng Yang, Steven Franconeri, Timothy F. Brady
- Visual Thinking Lab @ Northwestern University et al.
- Featured at IEEE VIS 2021

arXiv:2107.11367v1 [cs.HC] 23 Jul 2021

Rethinking the Ranks of Visual Channels

Caitlyn M. McColeman*, Fumeng Yang*, Timothy F. Brady, and Steven Franconeri

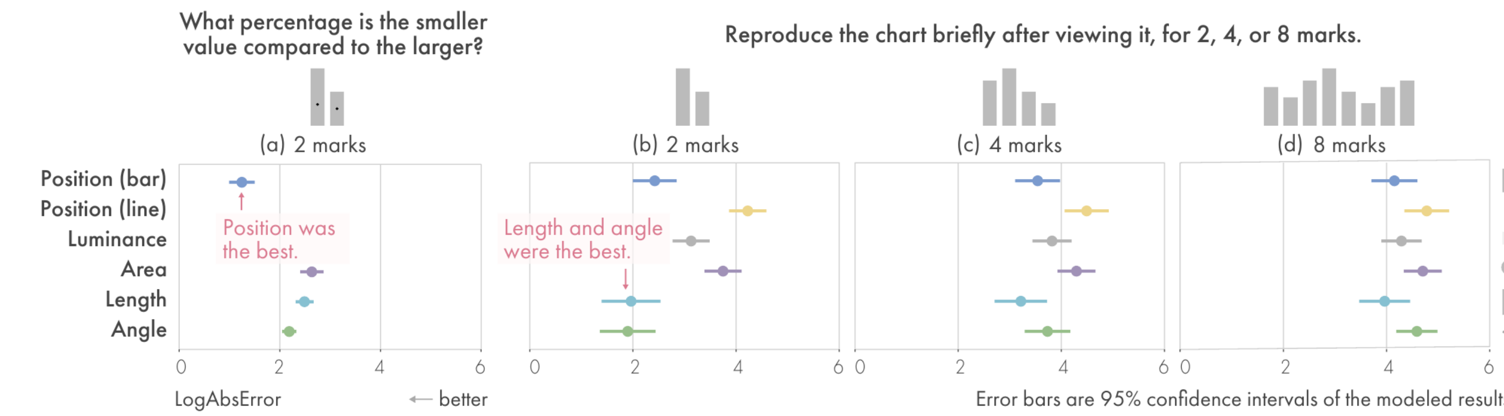


Fig. 1. One core guideline for data visualization design is that some visual channels offer better perceptual precision than others, drawing those precision estimates from two-value ratio judgment tasks [17]. (a) This figure depicts typical data (from [33], 50 participants) showing these judgments are more precise for position (e.g., bar graphs) than for area (e.g., bubble charts). We tested whether that ranking generalizes to the new task of reproducing 2 to 8 previously seen values, and analyzed reproduction bias, precision, and error using a Bayesian modeling approach. (b) This figure shows our modeled results (49 participants). The ranking did not hold, and other factors besides channel choice—like the number of values in the series—had an order of magnitude more influence on performance.

Abstract—Data can be visually represented using visual channels like position, length or luminance. An existing ranking of these visual channels is based on how accurately participants could report the ratio between two depicted values. There is an assumption that this ranking should hold for different tasks and for different numbers of marks. However, there is surprisingly little existing work that tests this assumption, especially given that visually computing ratios is relatively unimportant in real-world visualizations, compared to seeing, remembering, and comparing trends and motifs, across displays that almost universally depict more than two values.
To simulate the information extracted from a glance at a visualization, we instead asked participants to immediately reproduce a set of values from memory after they were shown the visualization. These values could be shown in a bar graph (position (bar)), line graph (position (line)), heat map (luminance), bubble chart (area), misaligned bar graph (length), or 'wind map' (angle). With a Bayesian multilevel modeling approach, we show how the rank positions of visual channels shift across different numbers of marks (2, 4 or 8) and for bias, precision, and error measures. The ranking did not hold, even for reproductions of only 2 marks, and the new probabilistic ranking was highly inconsistent for reproductions of different numbers of marks. Other factors besides channel choice had an order of magnitude more influence on performance, such as the number of values in the series (e.g., more marks led to larger errors), or the value of each mark (e.g., small values were systematically overestimated). Every visual channel was worse for displays with 8 marks than 4, consistent with established limits on visual memory. These results point to the need for a body of empirical studies that move beyond two-value ratio judgments as a baseline for reliably ranking the quality of a visual channel, including testing new tasks (detection of trends or motifs), timescales (immediate computation, or later comparison), and the number of values (from a handful, to thousands).

Index Terms—DataType Agnostic; Human-Subjects Quantitative Studies; Perception & Cognition; Charts, Diagrams, and Plots.

1 Introduction

Metric values can be efficiently transmitted to the human visual system across a set of channels, including position, length, or intensity [6] (see [56] for review). When creating a visualization, designers face a choice of which channel to depict metric values, with a major constraint being a ranking of putative *perceptual precision* of that channel. This ranking is organized by either expert judgment [50] or operationalized by a particular task. The most referenced operationalization is the precision of making ratio judgments between two values [17,30,33,77].

For example, in Fig. 2a, the viewer might use the position channel to estimate that the value for A is 85% of the value for B, close to the correct answer of 80%. The typical (log) error for this judgment is shown in Fig. 1a. It is typically the lowest error of any channel. Making the same judgment in Fig. 2b between A and B (now separated vertically) is a bit tougher, as reflected by the larger error value for ratio judgments of length in Fig. 1a. Finally, Fig. 2c shows the same data plotted as luminances. While we do not know of empirical measurements of ratio judgment error for this channel, expert judgment [17] (as well as ours) suggests that the error would be quite high [50].

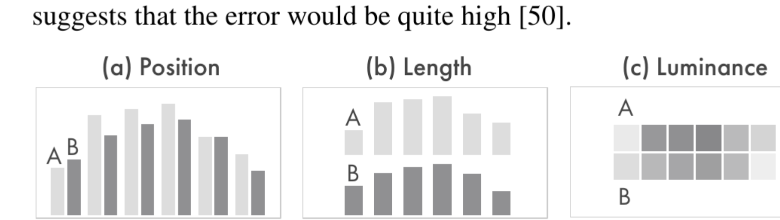


Fig. 2. **Examples of visualization designs that use three different visual channels.** (a) This bar chart relies on the position channel for comparison, (b) this bar chart relies on the length channel for vertical comparisons between A and B, and (c) this heatmap relies on the luminance channel. A two-value ratio judgment is precise in (a), and progressively less precise from (b) to (c).

- *Caitlyn McColeman* is with Northwestern University. E-mail: caitlyn.mccoleman@gmail.com.*
- *Fumeng Yang* is with Brown University. E-mail:fy@brown.edu.*
  ** Both authors contributed equally to this work.*
- *Timothy Brady is with University of San Diego. E-mail:timbrady@ucsd.edu.*
- *Steven Franconeri is with Northwestern University. E-mail:franconeri@northwestern.edu.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx/

Rethinking the Ranks of Visual Channels

PAPER

- Caitlyn M. McColeman, Fumeng Yang, Steven Franconeri, Timothy F. Brady
- Visual Thinking Lab @ Northwestern University et al.
- Featured at IEEE VIS 2021

CONTRIBUTIONS

1. Two experiments asking users to reproduce values
2. Data analysis via Bayesian hierarchical modeling
3. Probabilistic rankings for three metrics

arXiv:2107.11367v1 [cs.HC] 23 Jul 2021

Rethinking the Ranks of Visual Channels

Caitlyn M. McColeman*, Fumeng Yang*, Timothy F. Brady, and Steven Franconeri

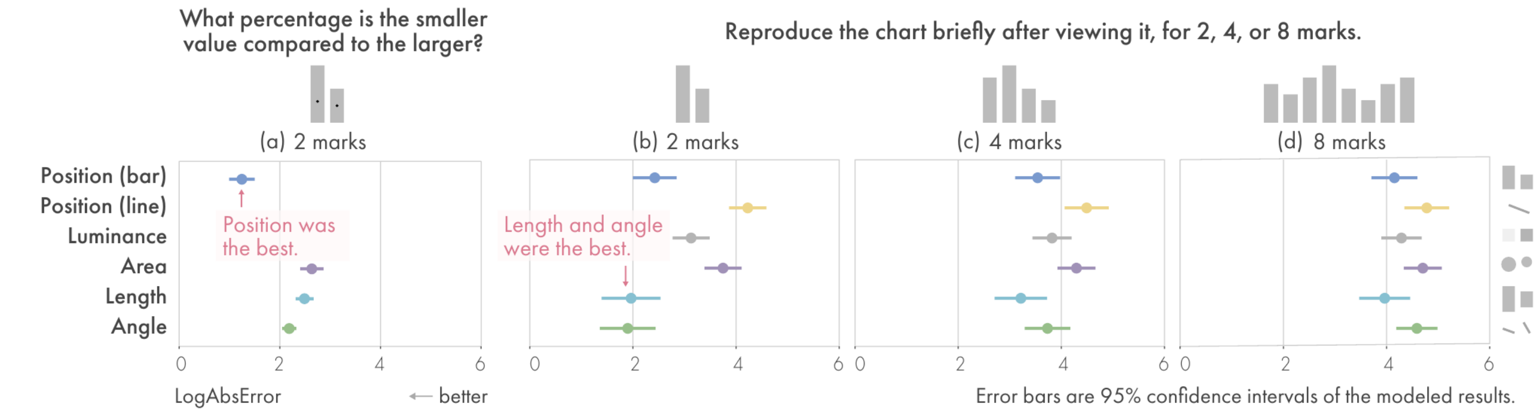


Fig. 1. One core guideline for data visualization design is that some visual channels offer better perceptual precision than others, drawing those precision estimates from two-value ratio judgment tasks [17]. (a) This figure depicts typical data (from [33], 50 participants) showing these judgments are more precise for position (e.g., bar graphs) than for area (e.g., bubble charts). We tested whether that ranking generalizes to the new task of reproducing 2 to 8 previously seen values, and analyzed reproduction bias, precision, and error using a Bayesian modeling approach. (b) This figure shows our modeled results (49 participants). The ranking did not hold, and other factors besides channel choice—like the number of values in the series—had an order of magnitude more influence on performance.

Abstract—Data can be visually represented using visual channels like position, length or luminance. An existing ranking of these visual channels is based on how accurately participants could report the ratio between two depicted values. There is an assumption that this ranking should hold for different tasks and for different numbers of marks. However, there is surprisingly little existing work that tests this assumption, especially given that visually computing ratios is relatively unimportant in real-world visualizations, compared to seeing, remembering, and comparing trends and motifs, across displays that almost universally depict more than two values.
To simulate the information extracted from a glance at a visualization, we instead asked participants to immediately reproduce a set of values from memory after they were shown the visualization. These values could be shown in a bar graph (position (bar)), line graph (position (line)), heat map (luminance), bubble chart (area), misaligned bar graph (length), or 'wind map' (angle). With a Bayesian multilevel modeling approach, we show how the rank positions of visual channels shift across different numbers of marks (2, 4 or 8) and for bias, precision, and error measures. The ranking did not hold, even for reproductions of only 2 marks, and the new probabilistic ranking was highly inconsistent for reproductions of different numbers of marks. Other factors besides channel choice had an order of magnitude more influence on performance, such as the number of values in the series (e.g., more marks led to larger errors), or the value of each mark (e.g., small values were systematically overestimated). Every visual channel was worse for displays with 8 marks than 4, consistent with established limits on visual memory. These results point to the need for a body of empirical studies that move beyond two-value ratio judgments as a baseline for reliably ranking the quality of a visual channel, including testing new tasks (detection of trends or motifs), timescales (immediate computation, or later comparison), and the number of values (from a handful, to thousands).

Index Terms—DataType Agnostic; Human-Subjects Quantitative Studies; Perception & Cognition; Charts, Diagrams, and Plots.

1 Introduction

Metric values can be efficiently transmitted to the human visual system across a set of channels, including position, length, or intensity [6] (see [56] for review). When creating a visualization, designers face a choice of which channel to depict metric values, with a major constraint being a ranking of putative *perceptual precision* of that channel. This ranking is organized by either expert judgment [50] or operationalized by a particular task. The most referenced operationalization is the precision of making ratio judgments between two values [17,30,33,77].

For example, in Fig. 2a, the viewer might use the position channel to estimate that the value for A is 85% of the value for B, close to the correct answer of 80%. The typical (log) error for this judgment is shown in Fig. 1a. It is typically the lowest error of any channel. Making the same judgment in Fig. 2b between A and B (now separated vertically) is a bit tougher, as reflected by the larger error value for ratio judgments of length in Fig. 1a. Finally, Fig. 2c shows the same data plotted as luminances. While we do not know of empirical measurements of ratio judgment error for this channel, expert judgment [17] (as well as ours) suggests that the error would be quite high [50].

- *Caitlyn McColeman* is with Northwestern University. E-mail: caitlyn.mccoleman@gmail.com.*
- *Fumeng Yang* is with Brown University. E-mail:fy@brown.edu.* * *Both authors contributed equally to this work.*
- *Timothy Brady is with University of San Diego. E-mail:timbrady@ucsd.edu.*
- *Steven Franconeri is with Northwestern University. E-mail:franconeri@northwestern.edu.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org.
Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx/

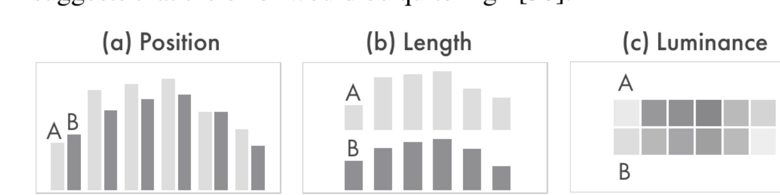


Fig. 2. **Examples of visualization designs that use three different visual channels.** (a) This bar chart relies on the position channel for comparison, (b) this bar chart relies on the length channel for vertical comparisons between A and B, and (c) this heatmap relies on the luminance channel. A two-value ratio judgment is precise in (a), and progressively less precise from (b) to (c).

Rethinking the Ranks of Visual Channels

PAPER

- Caitlyn M. McColeman, Fumeng Yang, Steven Franconeri, Timothy F. Brady
- Visual Thinking Lab @ Northwestern University et al.
- Featured at IEEE VIS 2021

CONTRIBUTIONS

1. Two experiments asking users to reproduce values
2. Data analysis via Bayesian hierarchical modeling
3. Probabilistic rankings for three metrics

And more questions

arXiv:2107.11367v1 [cs.HC] 23 Jul 2021

Rethinking the Ranks of Visual Channels

Caitlyn M. McColeman*, Fumeng Yang*, Timothy F. Brady, and Steven Franconeri

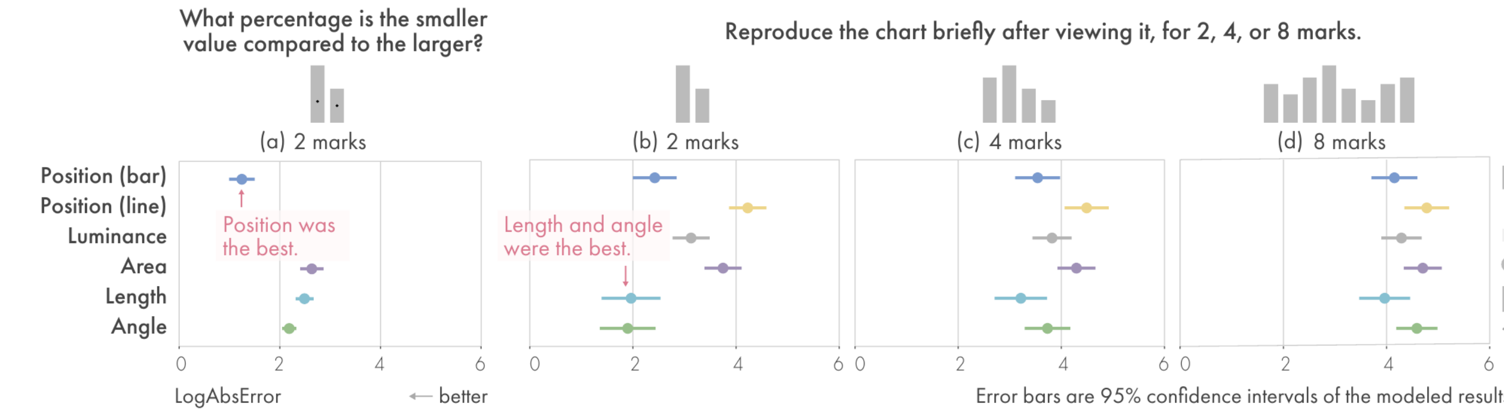


Fig. 1. One core guideline for data visualization design is that some visual channels offer better perceptual precision than others, drawing those precision estimates from two-value ratio judgment tasks [17]. (a) This figure depicts typical data (from [33], 50 participants) showing these judgments are more precise for position (e.g., bar graphs) than for area (e.g., bubble charts). We tested whether that ranking generalizes to the new task of reproducing 2 to 8 previously seen values, and analyzed reproduction bias, precision, and error using a Bayesian modeling approach. (b) This figure shows our modeled results (49 participants). The ranking did not hold, and other factors besides channel choice—like the number of values in the series—had an order of magnitude more influence on performance.

Abstract—Data can be visually represented using visual channels like position, length or luminance. An existing ranking of these visual channels is based on how accurately participants could report the ratio between two depicted values. There is an assumption that this ranking should hold for different tasks and for different numbers of marks. However, there is surprisingly little existing work that tests this assumption, especially given that visually computing ratios is relatively unimportant in real-world visualizations, compared to seeing, remembering, and comparing trends and motifs, across displays that almost universally depict more than two values.
To simulate the information extracted from a glance at a visualization, we instead asked participants to immediately reproduce a set of values from memory after they were shown the visualization. These values could be shown in a bar graph (position (bar)), line graph (position (line)), heat map (luminance), bubble chart (area), misaligned bar graph (length), or 'wind map' (angle). With a Bayesian multilevel modeling approach, we show how the rank positions of visual channels shift across different numbers of marks (2, 4 or 8) and for bias, precision, and error measures. The ranking did not hold, even for reproductions of only 2 marks, and the new probabilistic ranking was highly inconsistent for reproductions of different numbers of marks. Other factors besides channel choice had an order of magnitude more influence on performance, such as the number of values in the series (e.g., more marks led to larger errors), or the value of each mark (e.g., small values were systematically overestimated). Every visual channel was worse for displays with 8 marks than 4, consistent with established limits on visual memory. These results point to the need for a body of empirical studies that move beyond two-value ratio judgments as a baseline for reliably ranking the quality of a visual channel, including testing new tasks (detection of trends or motifs), timescales (immediate computation, or later comparison), and the number of values (from a handful, to thousands).

Index Terms—DataType Agnostic; Human-Subjects Quantitative Studies; Perception & Cognition; Charts, Diagrams, and Plots.

1 Introduction

Metric values can be visually efficiently transmitted to the human visual system across a set of channels, including position, length, or intensity [6] (see [56] for review). When creating a visualization, designers face a choice of which channel to depict metric values, with a major constraint being a ranking of putative *perceptual precision* of that channel. This ranking is organized by either expert judgment [50] or operationalized by a particular task. The most referenced operationalization is the precision of making ratio judgments between two values [17,30,33,77].

For example, in Fig. 2a, the viewer might use the position channel to estimate that the value for A is 85% of the value for B, close to the correct answer of 80%. The typical (log) error for this judgment is shown in Fig. 1a. It is typically the lowest error of any channel. Making the same judgment in Fig. 2b between A and B (now separated vertically) is a bit tougher, as reflected by the larger error value for ratio judgments of length in Fig. 1a. Finally, Fig. 2c shows the same data plotted as luminances. While we do not know of empirical measurements of ratio judgment error for this channel, expert judgment [17] (as well as ours) suggests that the error would be quite high [50].

- *Caitlyn McColeman* is with Northwestern University. E-mail: caitlyn.mccoleman@gmail.com.*
- *Fumeng Yang* is with Brown University. E-mail:fy@brown.edu.*
  ** Both authors contributed equally to this work.*
- *Timothy Brady is with University of San Diego. E-mail:timbrady@ucsd.edu.*
- *Steven Franconeri is with Northwestern University. E-mail:franconeri@northwestern.edu.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org.
Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx/

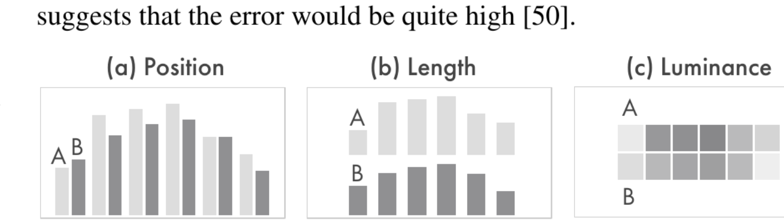


Fig. 2. **Examples of visualization designs that use three different visual channels.** (a) This bar chart relies on the position channel for comparison, (b) this bar chart relies on the length channel for vertical comparisons between A and B, and (c) this heatmap relies on the luminance channel. A two-value ratio judgment is precise in (a), and progressively less precise from (b) to (c).

Steve Franconeri
@SteveFranconeri

Replying to @domoritz @fumeng_yang and 2 others

'Consider' is a good term for the moment, this ranking is not ready for implementation. The paper challenges the existing ranking method, but more work is needed using more realistic tasks before generating concrete guidelines.

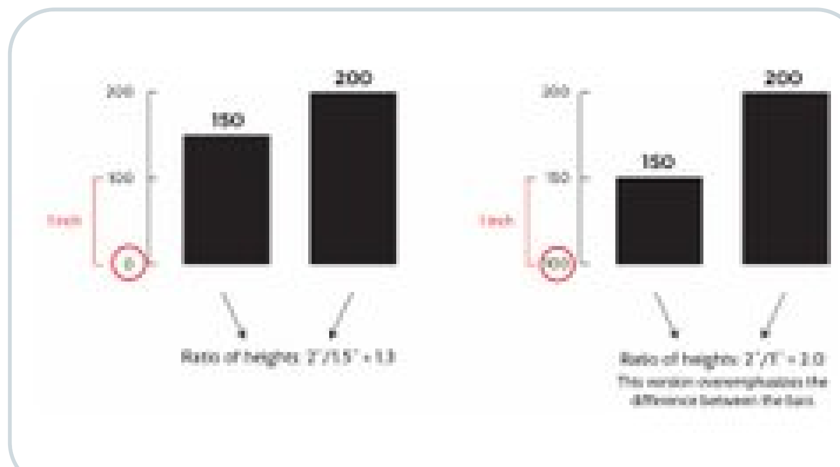
5:51 PM · Oct 23, 2021

4 Reply Cop...

Explore what's happening ...

Jon Schwabish
@jschwabish

I'm still on team 'bar charts start at zero' and here's an image that shows my mathematical reasoning. 📊 📊 📊



2:51 PM · Mar 10, 2022

90 Reply Co...

Read 2 replies

Steve Haroz 📊 👁 🧠
@sharoz

Replying to @jschwabish

If you want to take the "bar charts start at 0" stance, fine! But multiple papers have failed to find any difference between interpretation of bar charts and dot/line charts. So, you should have the same view there.

3:09 PM · Mar 10, 2022

2 Reply Cop...

Read 1 reply

Thanks!

Some chart types/visual encodings are more "precise", based on asking people to judge a ratio between two values. The authors asked people to reproduce up to 8 values. Cleveland & McGill's ranking did not hold. The rankings are distinct for a different task, metrics, and numbers of marks.