

Data Visualization

Becoming a data vizard



Montana Data Conference

September 15, 2016

Data visualization = Storytelling



Citation: Stephanie Evergreen

Blog: stephanieevergreen.com
annkemery.com

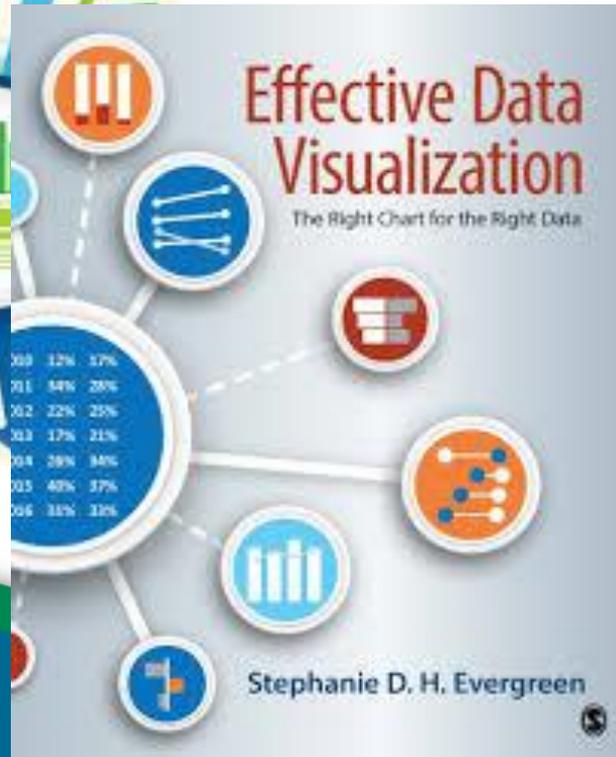
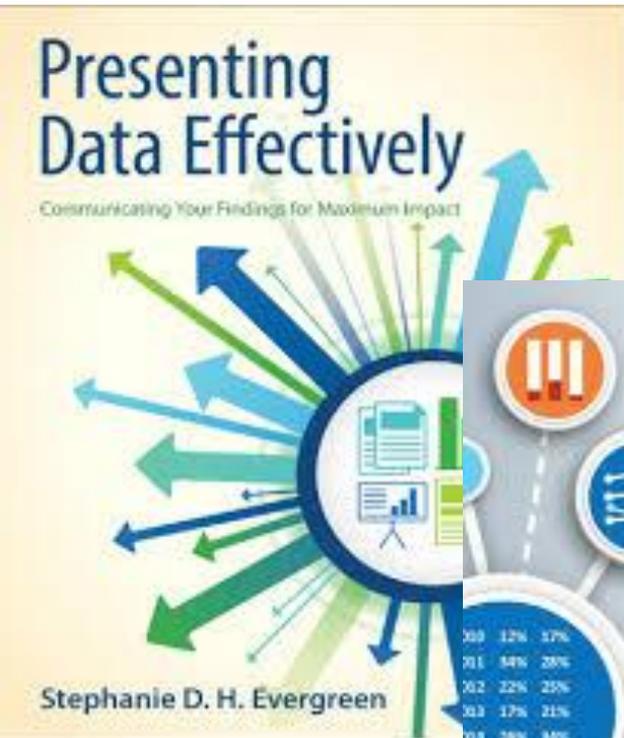
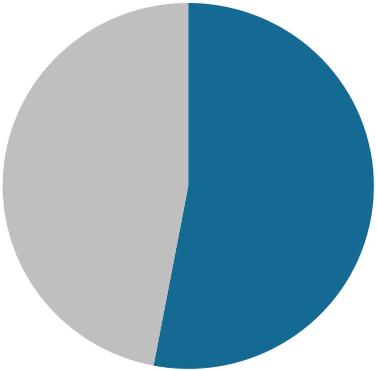
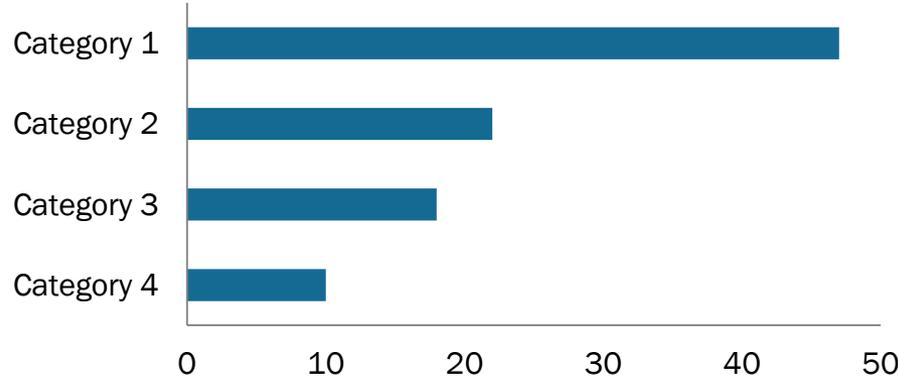


Chart types

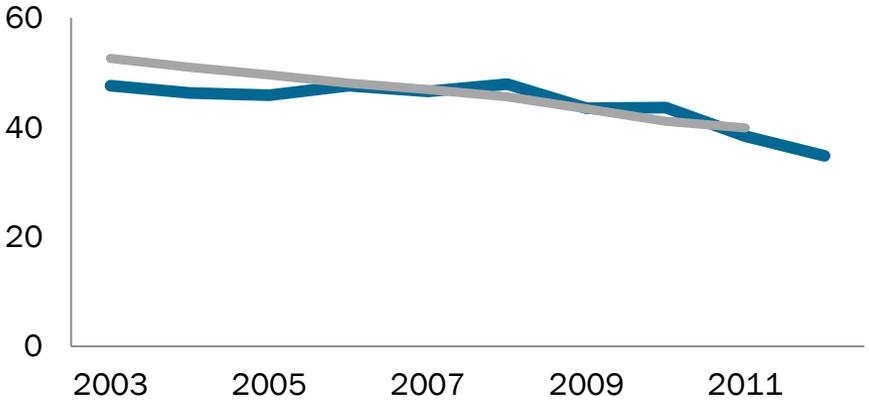
Pie Chart



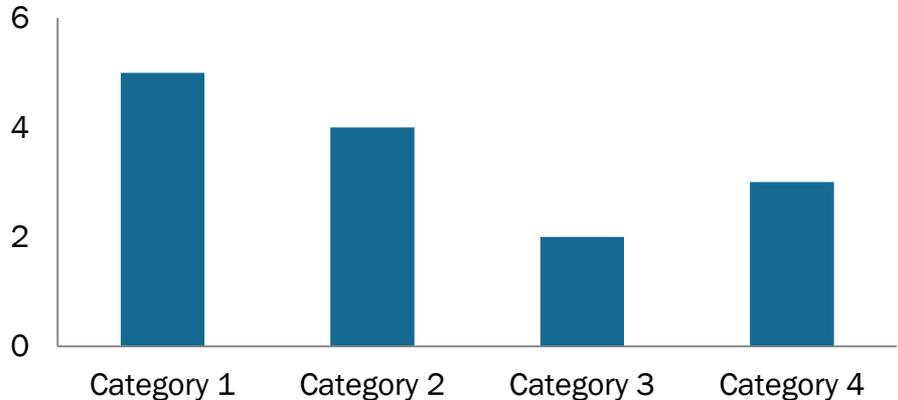
Bar Chart



Line Chart



Column chart



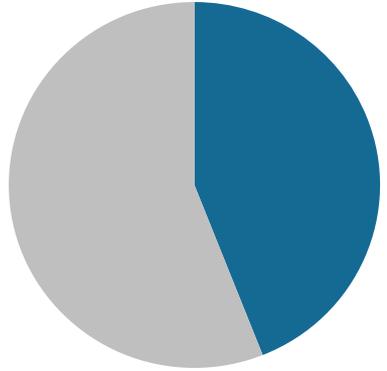
Emphasize one number

Big Number

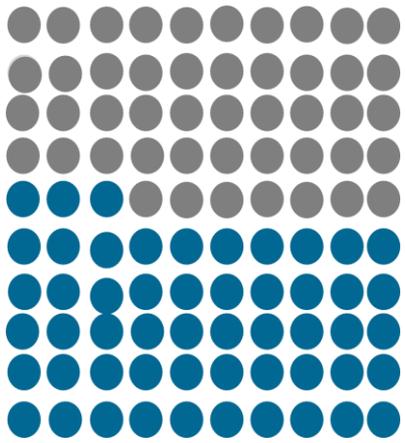
44%

“I didn’t think I needed one”

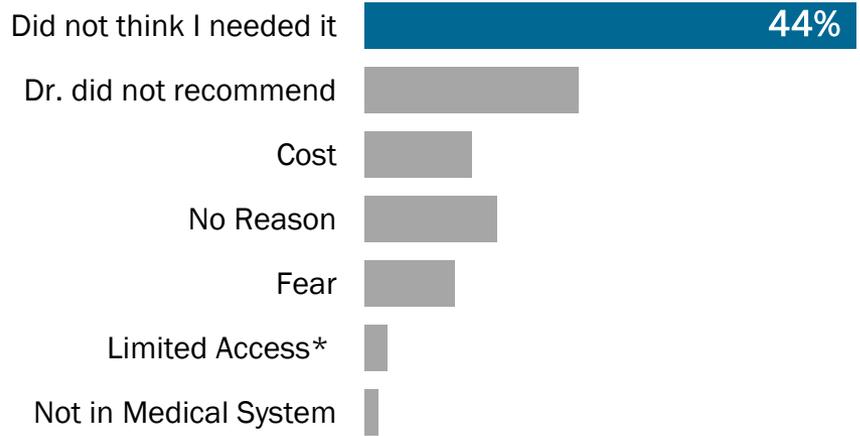
Pie Chart



Icon array

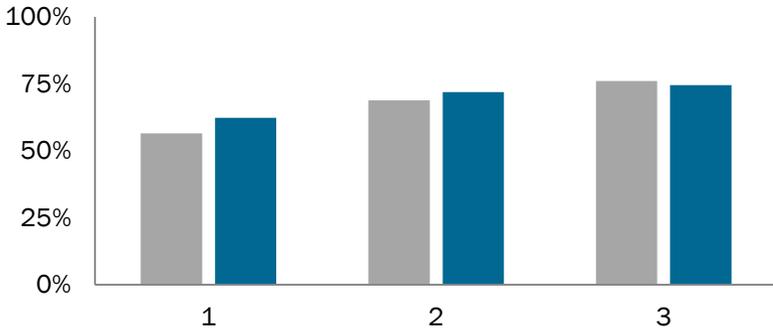


Bar Chart

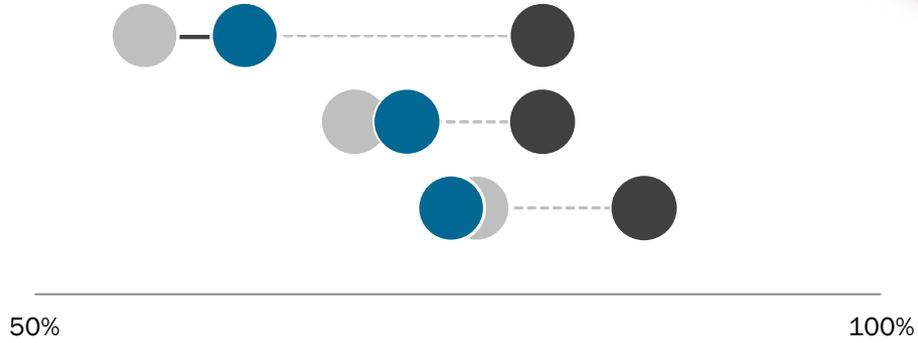


Comparing \geq two numbers

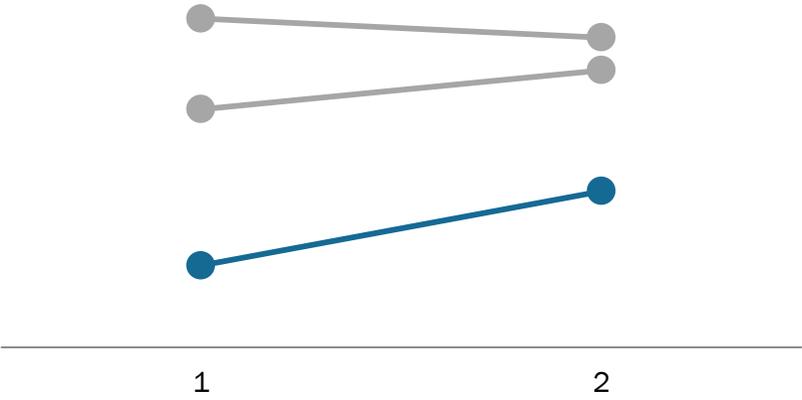
Side by Side



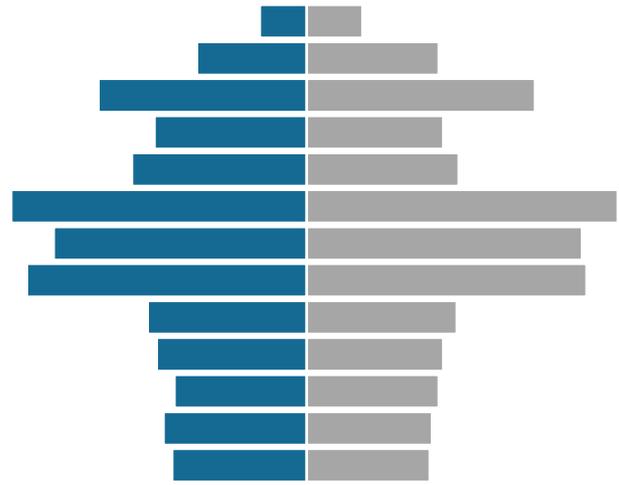
Dumbbell dot



Slope graph

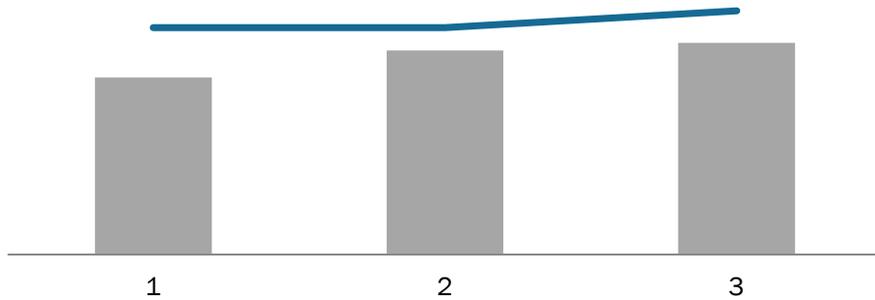


Back - to -Back



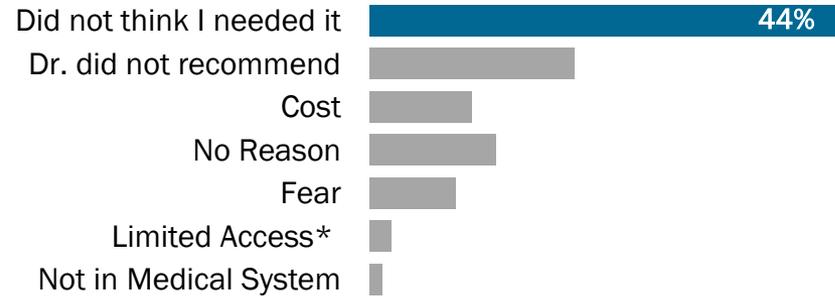
Compare to benchmark

Benchmark line

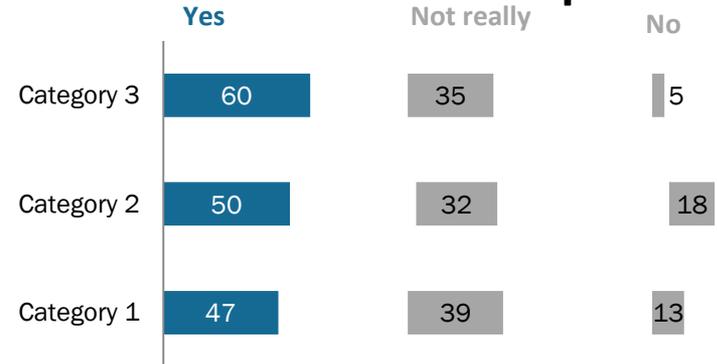


Survey data

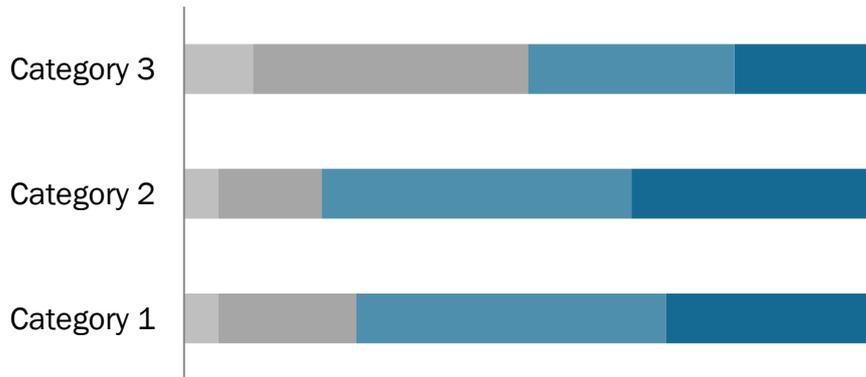
Bar Chart



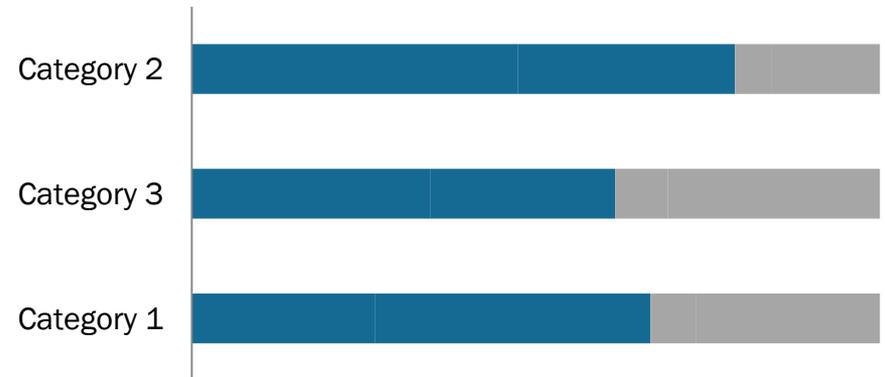
Small multiples



Stacked bar

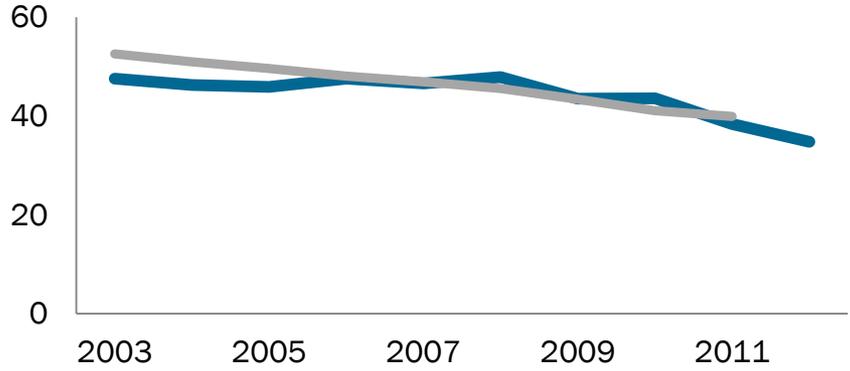


Aggregated bar

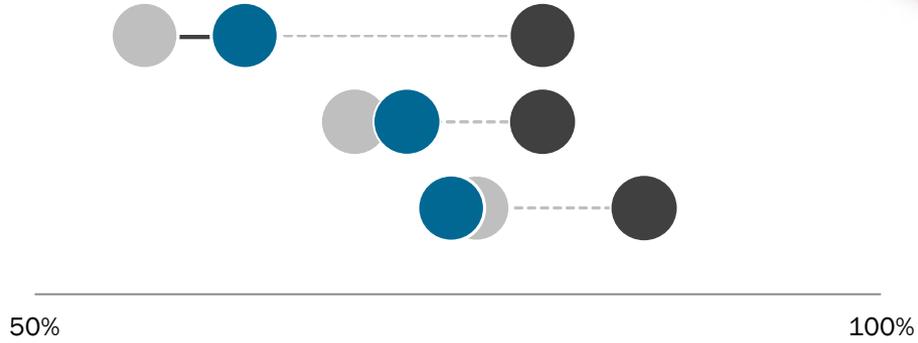


Change over time

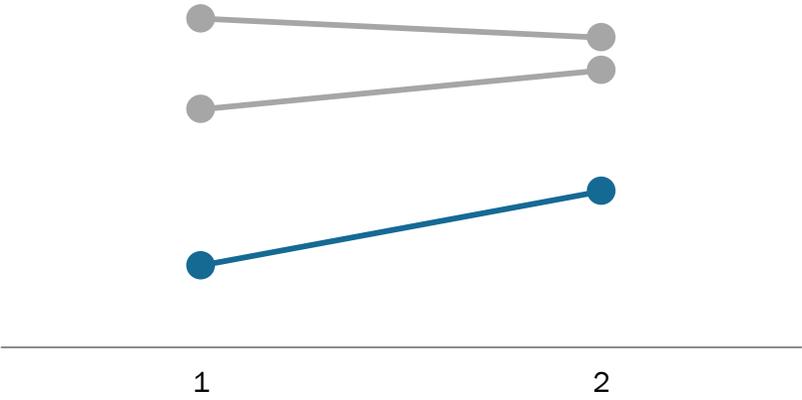
Line Chart



Dumbbell dot

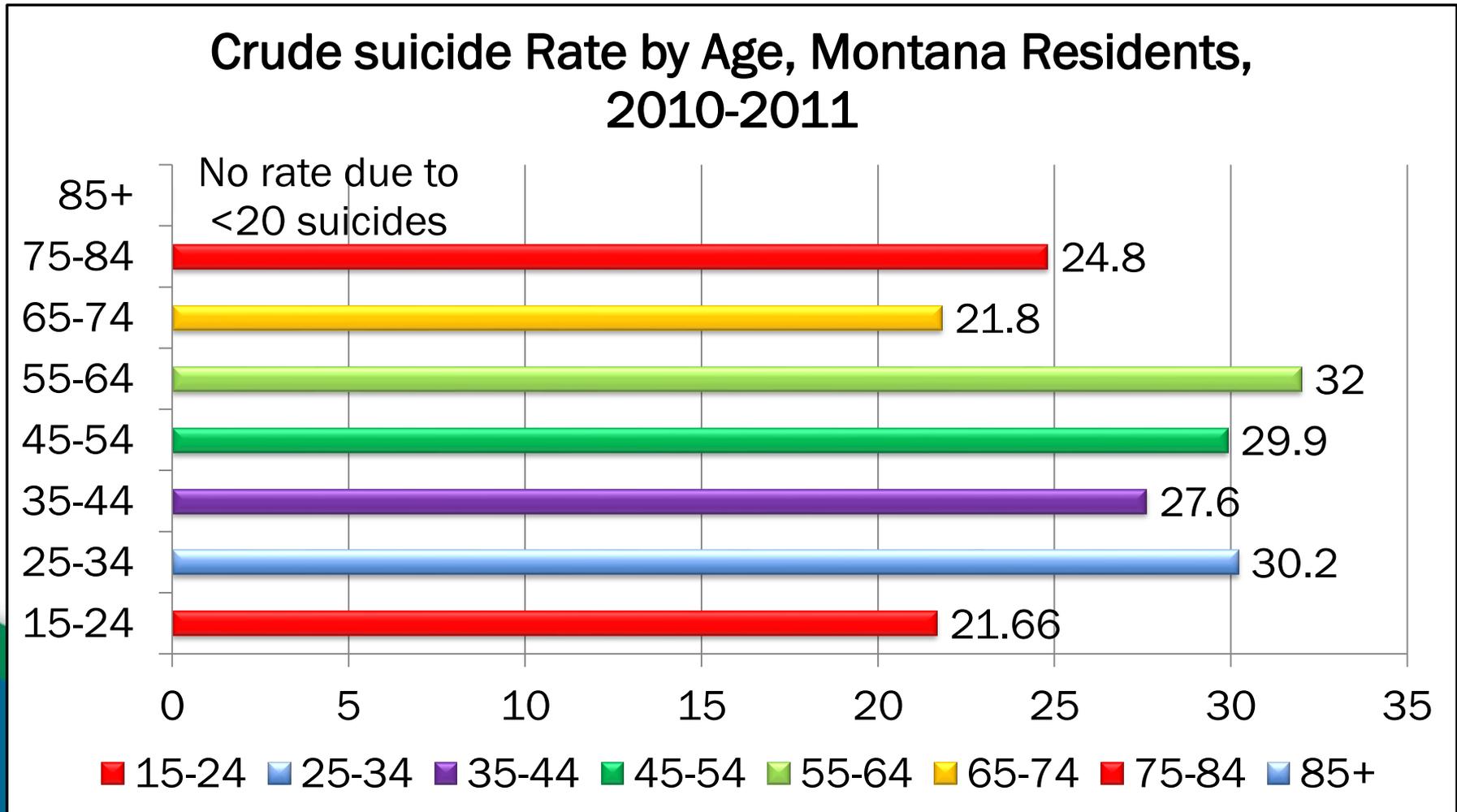


Slope graph



DESIGN PRINCIPLES

What message does this figure tell us?



OVERALL

Overall

Graphs will catch a viewer's attention so only visualize the data that needs attention. Too many graphics of unimportant information dilute the power of visualization.

Graph highlights significant finding or conclusion

Graphs should have a "so what?" – either a practical or statistical significance (or both) to warrant their presence.

The type of graph is appropriate for data

Data are displayed using a graph type appropriate for the relationship within the data. For example, change over time is displayed as a line graph, area chart, slope graph, or dot plot.

Graph has appropriate level of precision

Few numeric labels need decimal places. When precision is important, choose a type of graph type that displays differences through length or points along a line (e.g., bar charts, dot plots). When precision is less important, you can use a graph that displays differences through angles or area (e.g., pie charts, circle charts).

Contextualized or comparison data are present

Comparisons—over time, across programs or subgroups of participants, etc.—help the viewer understand the significance of the data.

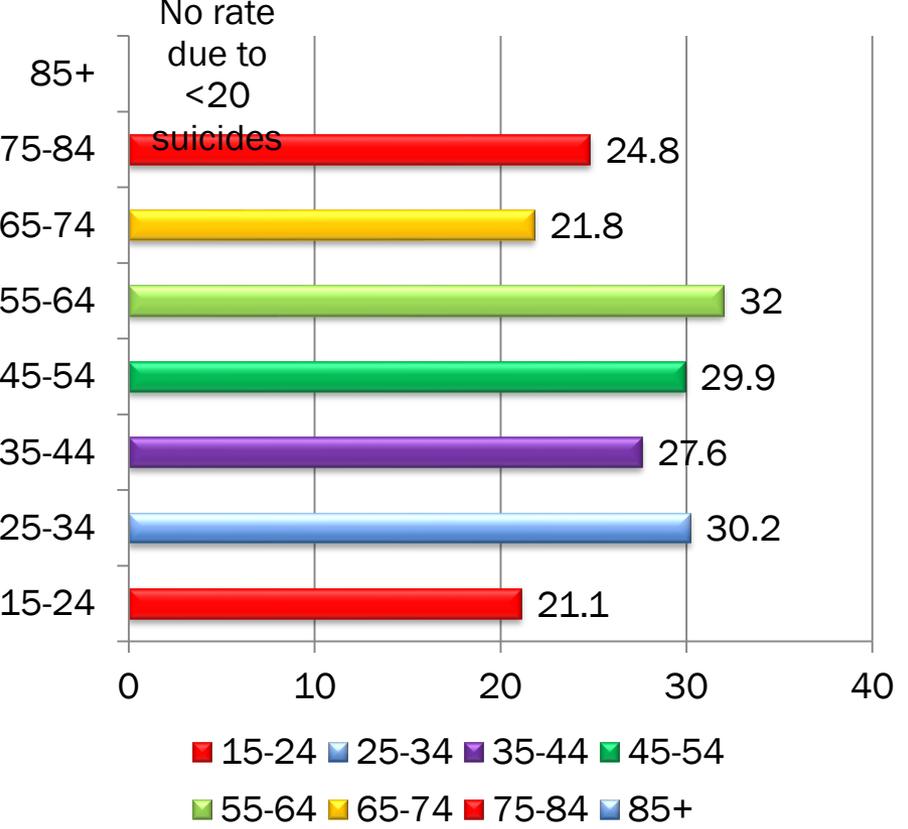
Individual chart elements work together to reinforce the overarching takeaway message

Choices about graph type, text, arrangement, color, and lines should reinforce the same takeaway message.

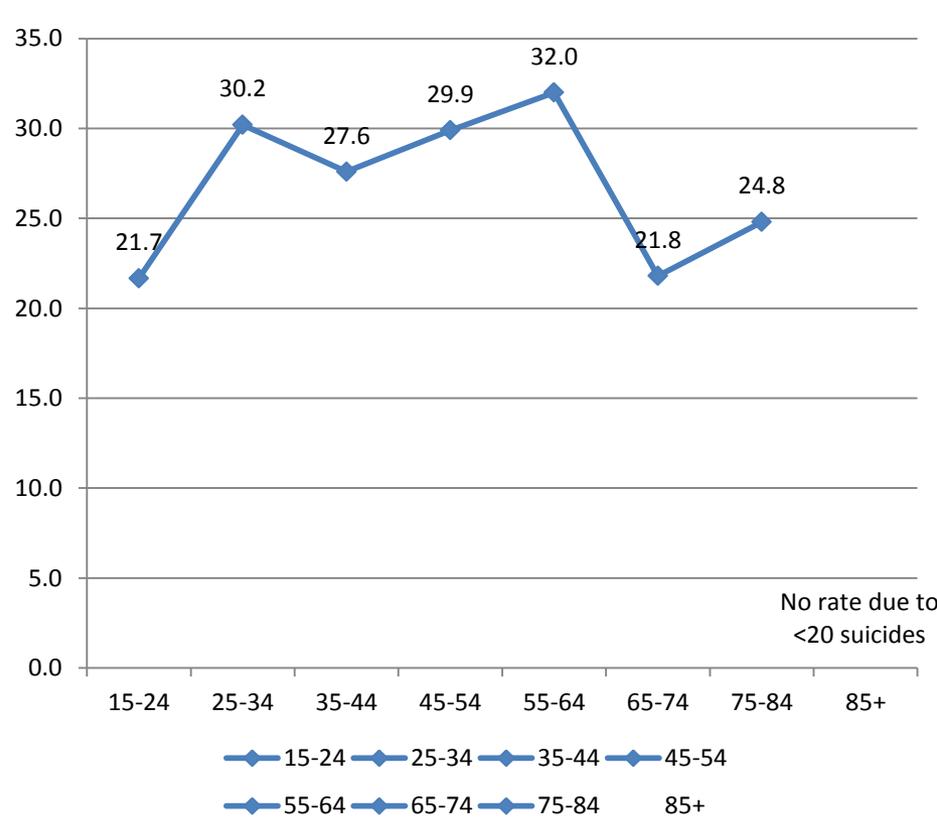
OVERALL:

Graph appropriate for data

Crude suicide Rate by Age, Montana Residents, 2010-2011



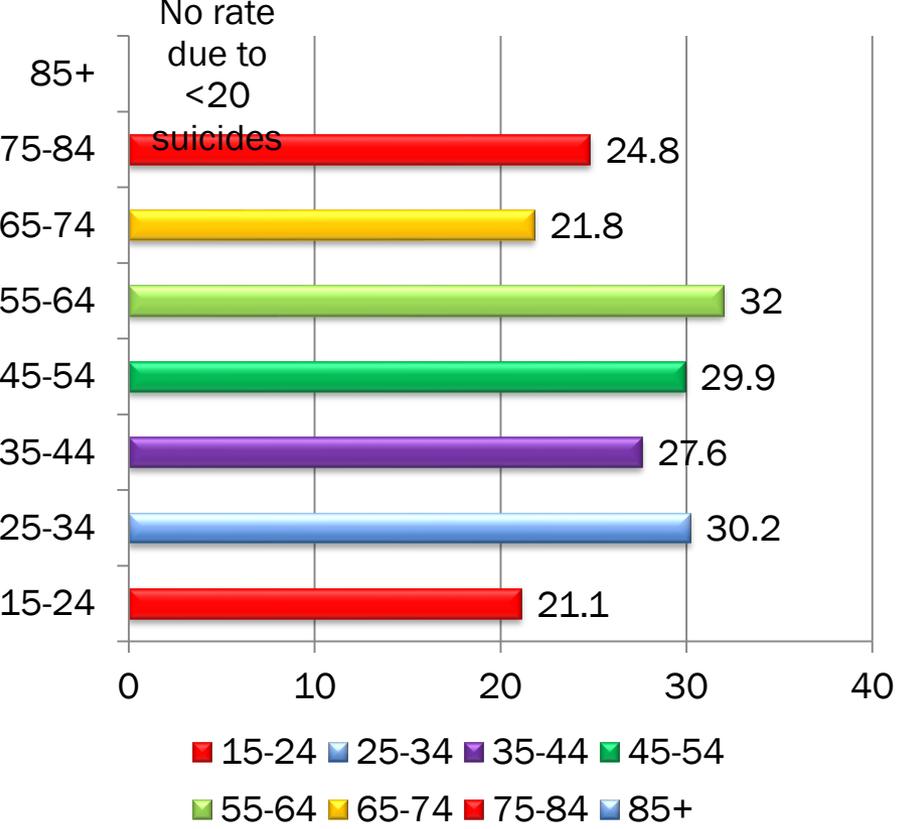
Crude suicide Rate by Age, Montana Residents, 2010-2011



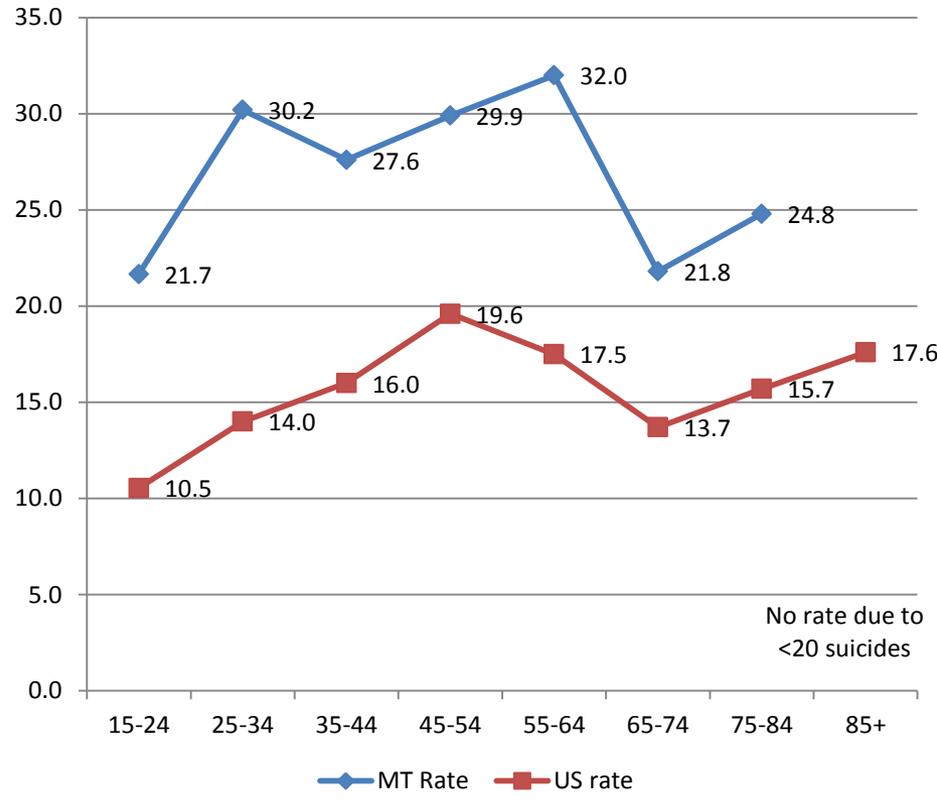
OVERALL:

Comparison data are present

Crude suicide Rate by Age, Montana Residents, 2010-2011



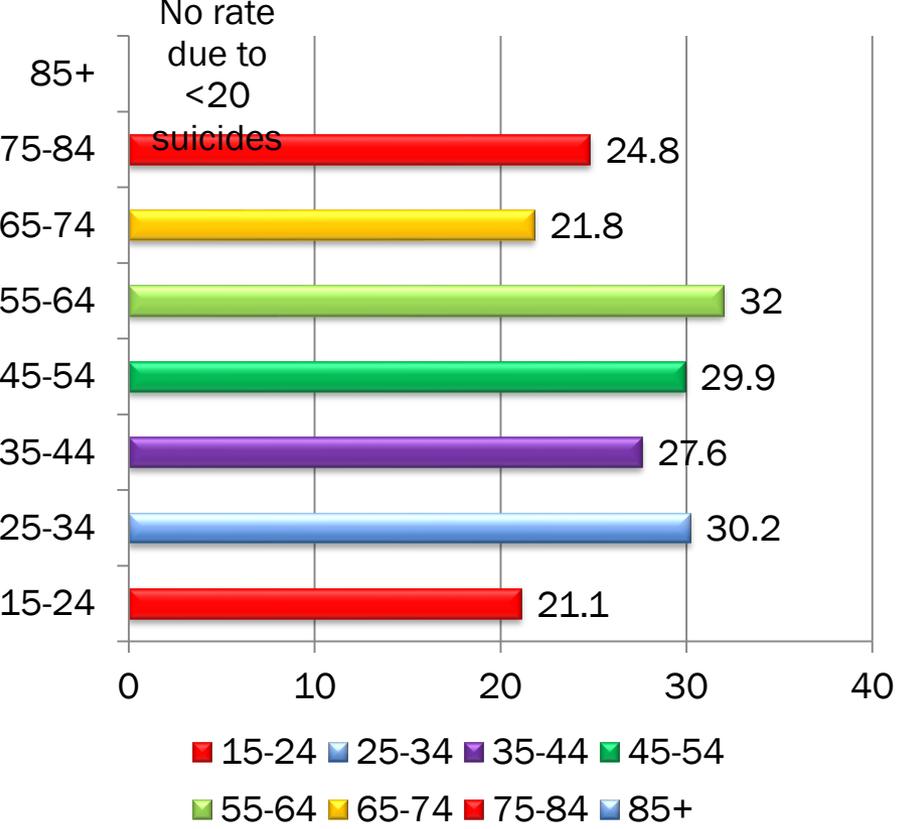
Crude suicide Rate by Age, Montana Residents, 2010-2011



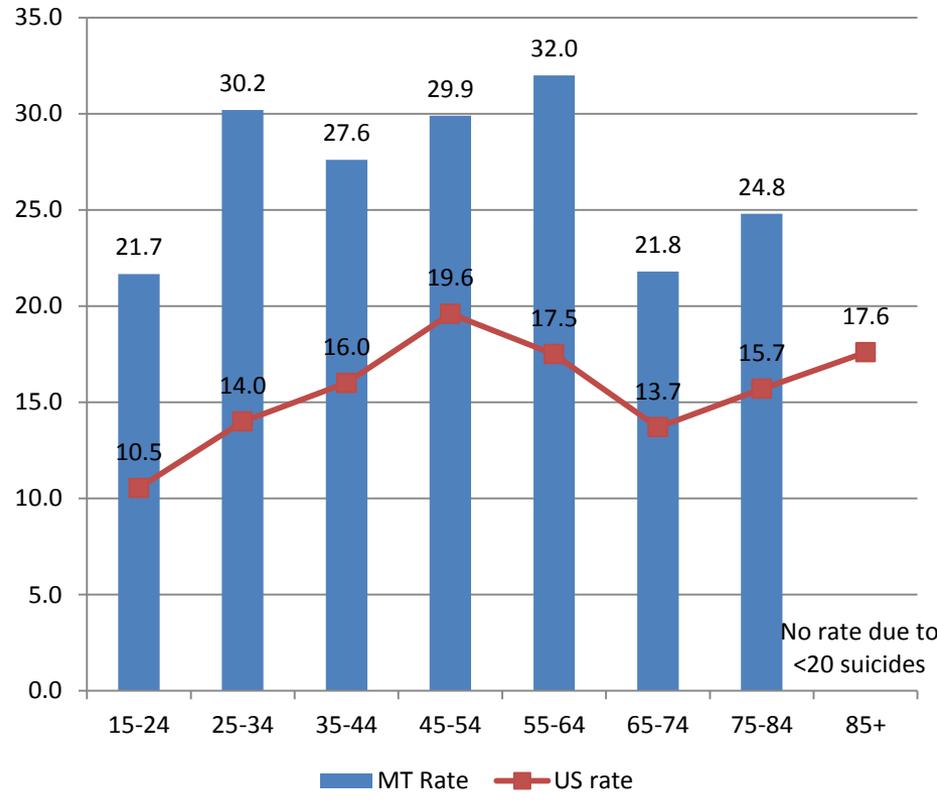
OVERALL:

Highlight significant findings

Crude suicide Rate by Age, Montana Residents, 2010-2011



Crude suicide Rate by Age, Montana Residents, 2010-2011



TEXT

Text

Graphs don't contain much text, so existing text must encapsulate your message and pack a punch.

6-12 word descriptive title is left-justified in upper left corner

Short titles enable readers to comprehend takeaway messages even while quickly skimming the graph. Rather than a generic phrase, use a descriptive sentence that encapsulates the graph's finding or "so what?" Western cultures start reading in the upper left, so locate the title there.

Subtitle and/or annotations provide additional information

Subtitles and annotations (call-out text within the graph) can add explanatory and interpretive power to a graph. Use them to answer questions a viewer might have or to highlight one or two data points.

Text size is hierarchical and readable

Titles are in a larger size than subtitles or annotations, which are larger than labels, which are larger than axis labels, which are larger than source information. The smallest text - axis labels - are at least 9 point font size on paper, at least 20 on screen.

Text is horizontal

Titles, subtitles, annotations, and data labels are horizontal (not vertical or diagonal). Line labels and axis labels can deviate from this rule and still receive full points.

Data are labeled directly

Position data labels near the data rather than in a separate legend (e.g., on top of or next to bars or pie slices, and next to lines in line charts). Eliminate/embed legends when possible because eye movement back and forth between the legend and the data can interrupt the brain's attempts to interpret the graph.

Labels are used sparingly

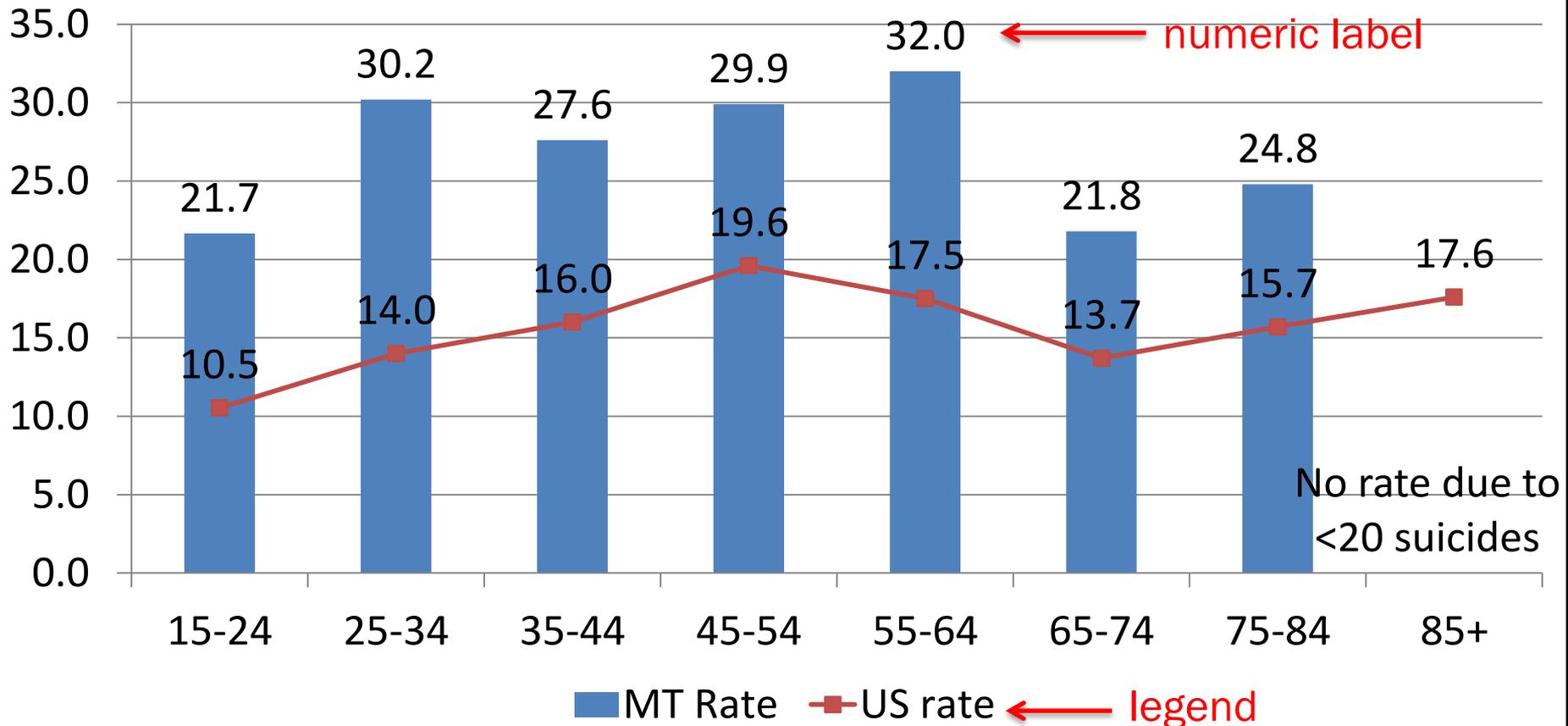
Focus attention by removing the redundancy. For example, in line charts, label every other year on an axis.

TEXT

title



Crude suicide Rate by Age, Montana Residents, 2010-2011



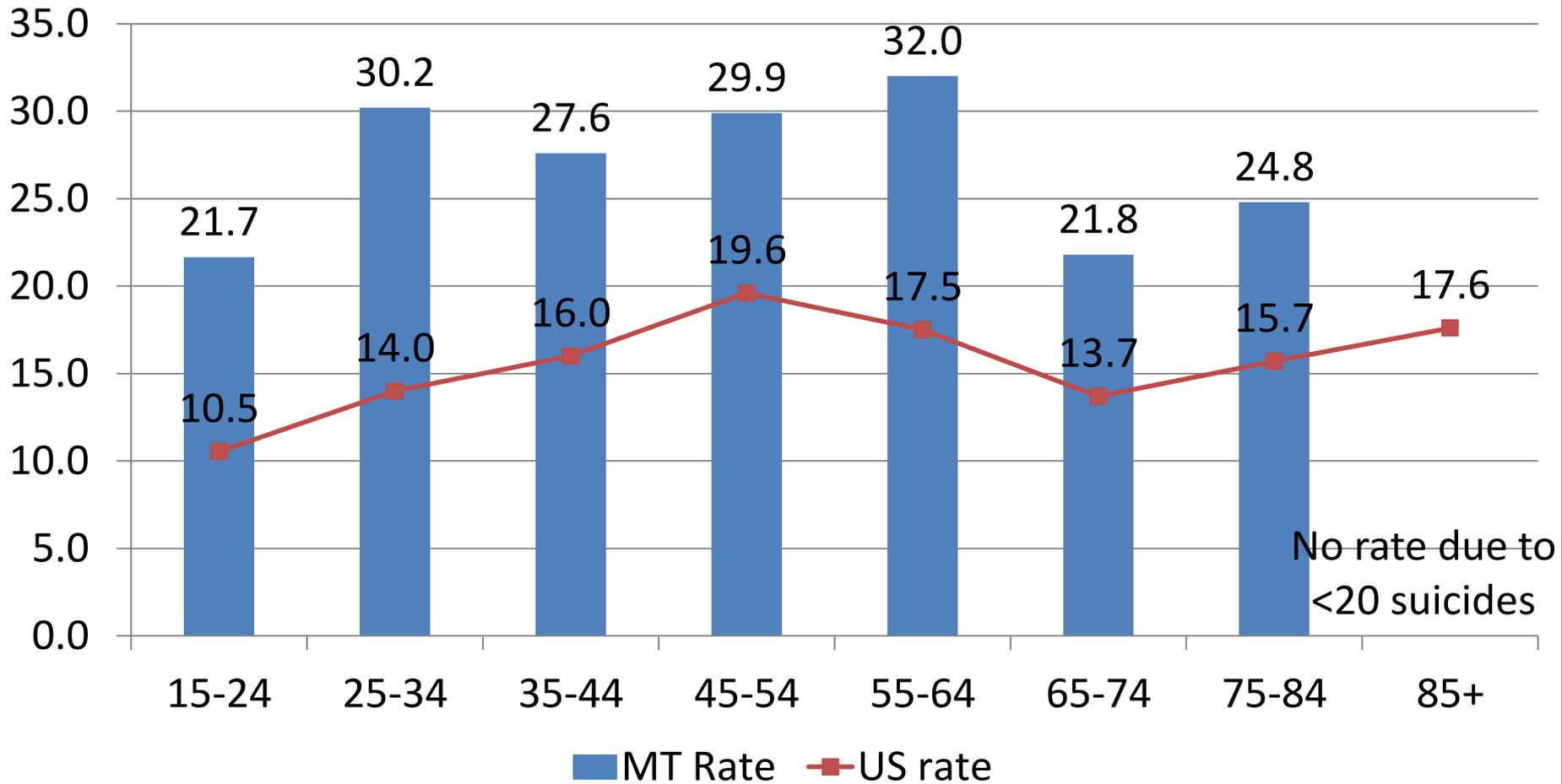
numeric label

legend

TEXT: Descriptive title



Suicide was higher in Montana compared to the U.S.

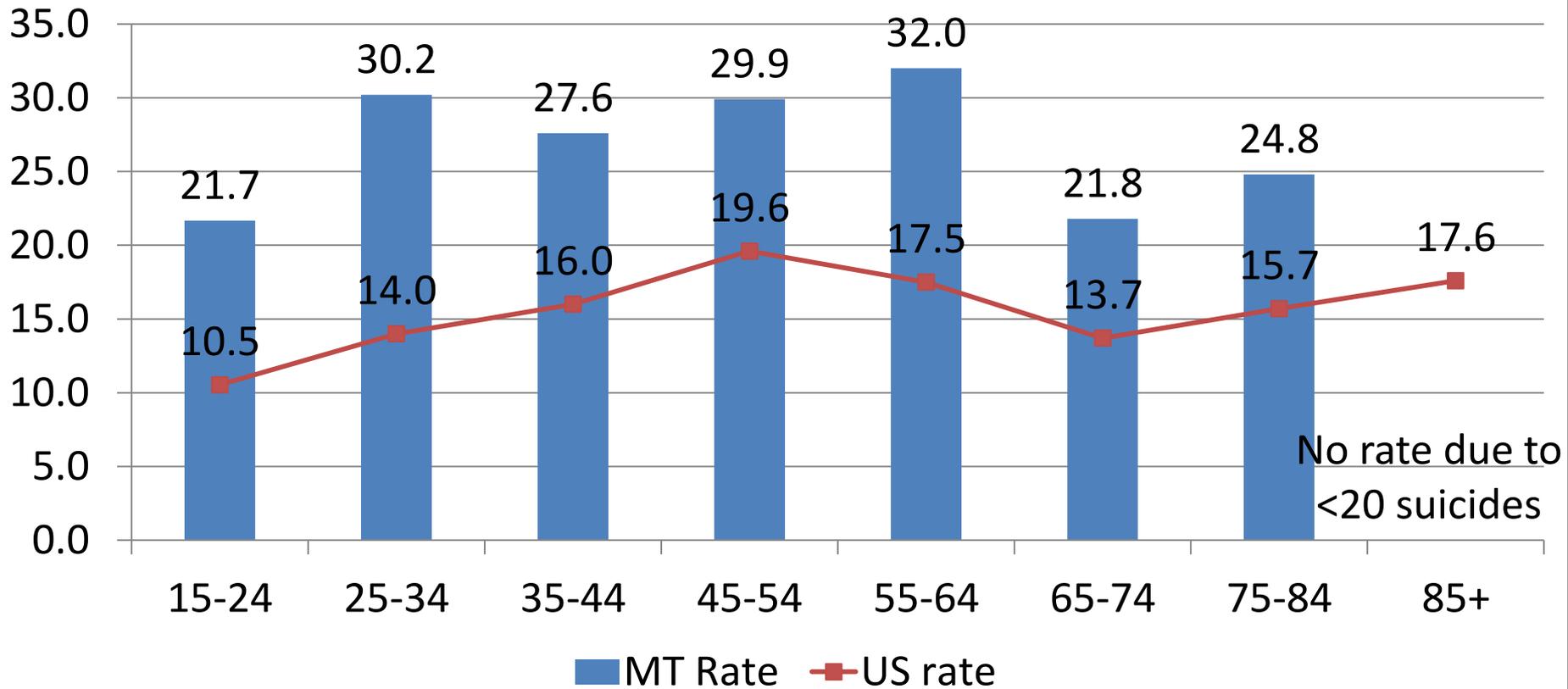


TEXT: Subtitle to provide additional information

Suicide was higher in Montana compared to the U.S.

In 2010, Montana suicide rates among youth (15-24) were two times higher than U.S.

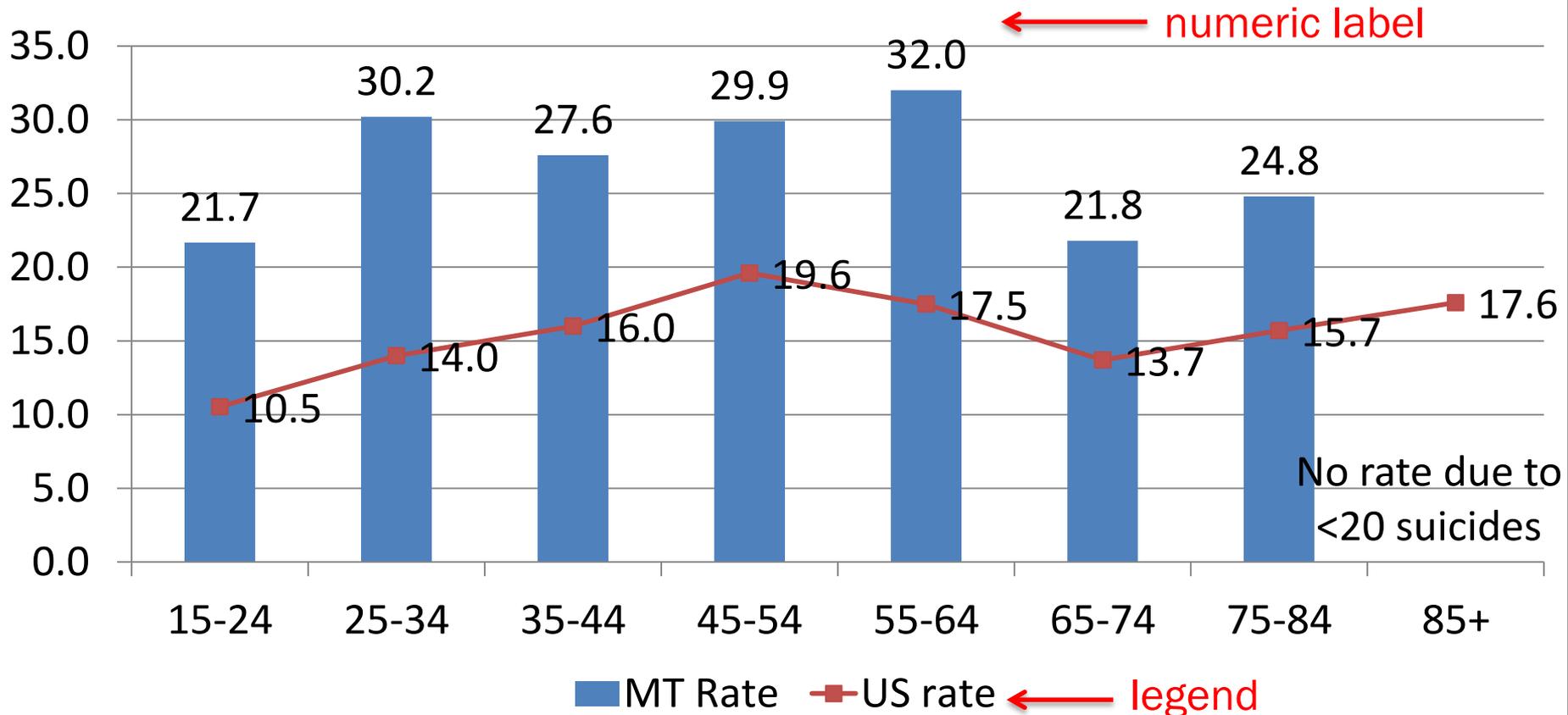
subtitle



TEXT: Directly label data, avoid legends

Suicide was higher in Montana compared to the U.S.

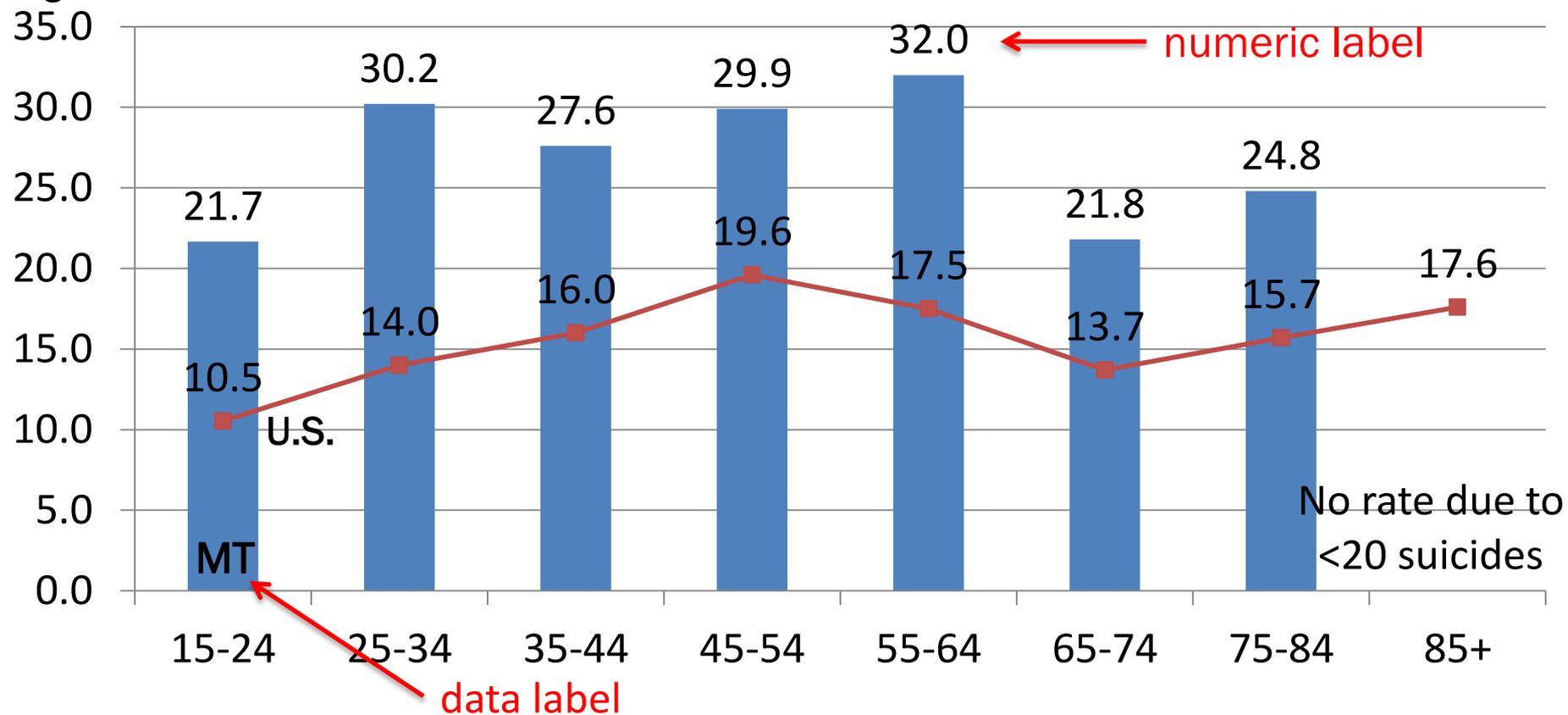
In 2010, Montana suicide rates among youth (15-24) were two times higher than U.S.



TEXT: Directly label data, avoid legends

Suicide was higher in Montana compared to the U.S.

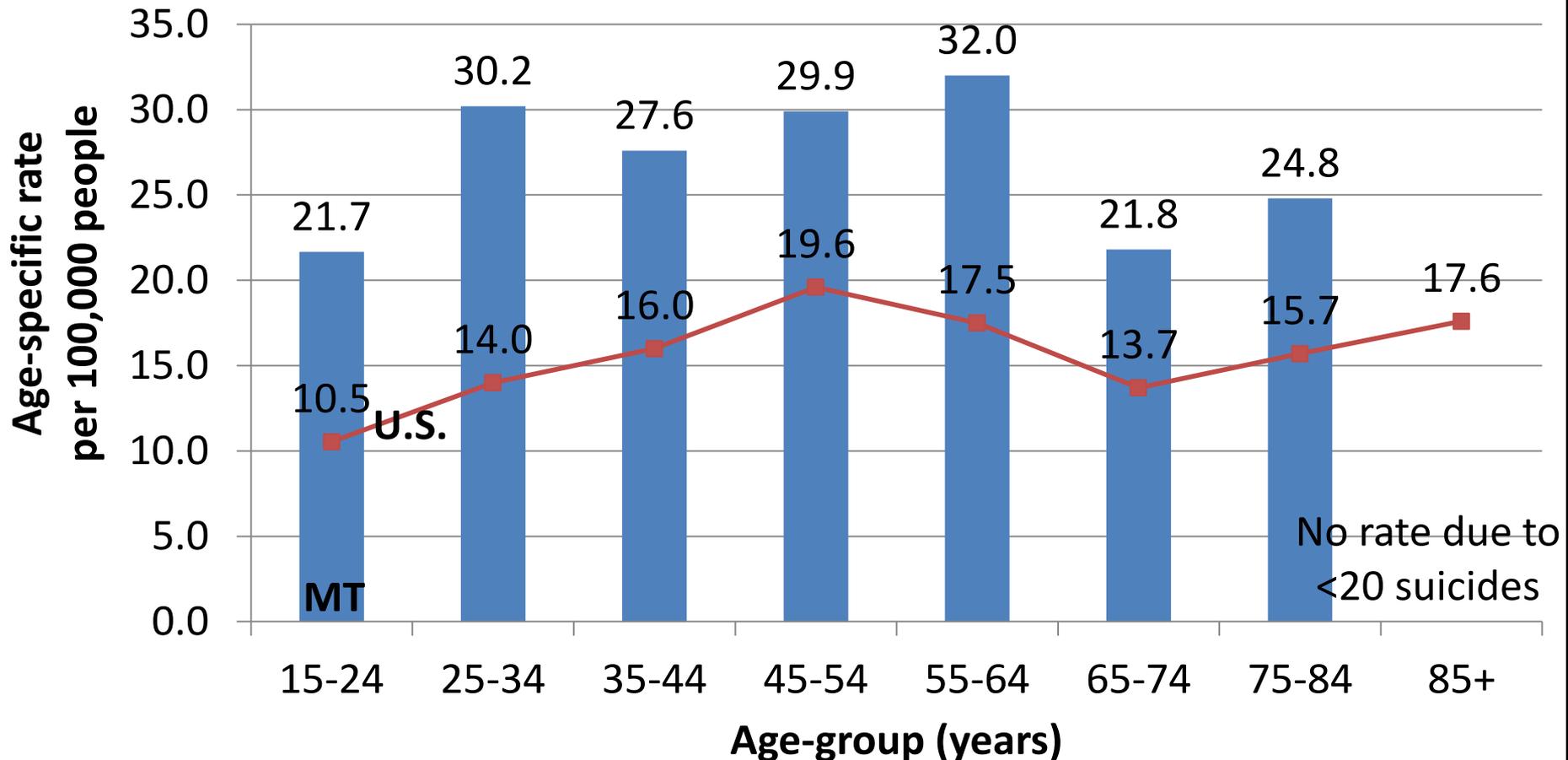
In 2010, Montana suicide rates among youth (15-24) were two times higher than U.S.



TEXT: Label axis

Suicide was higher in Montana compared to the U.S.

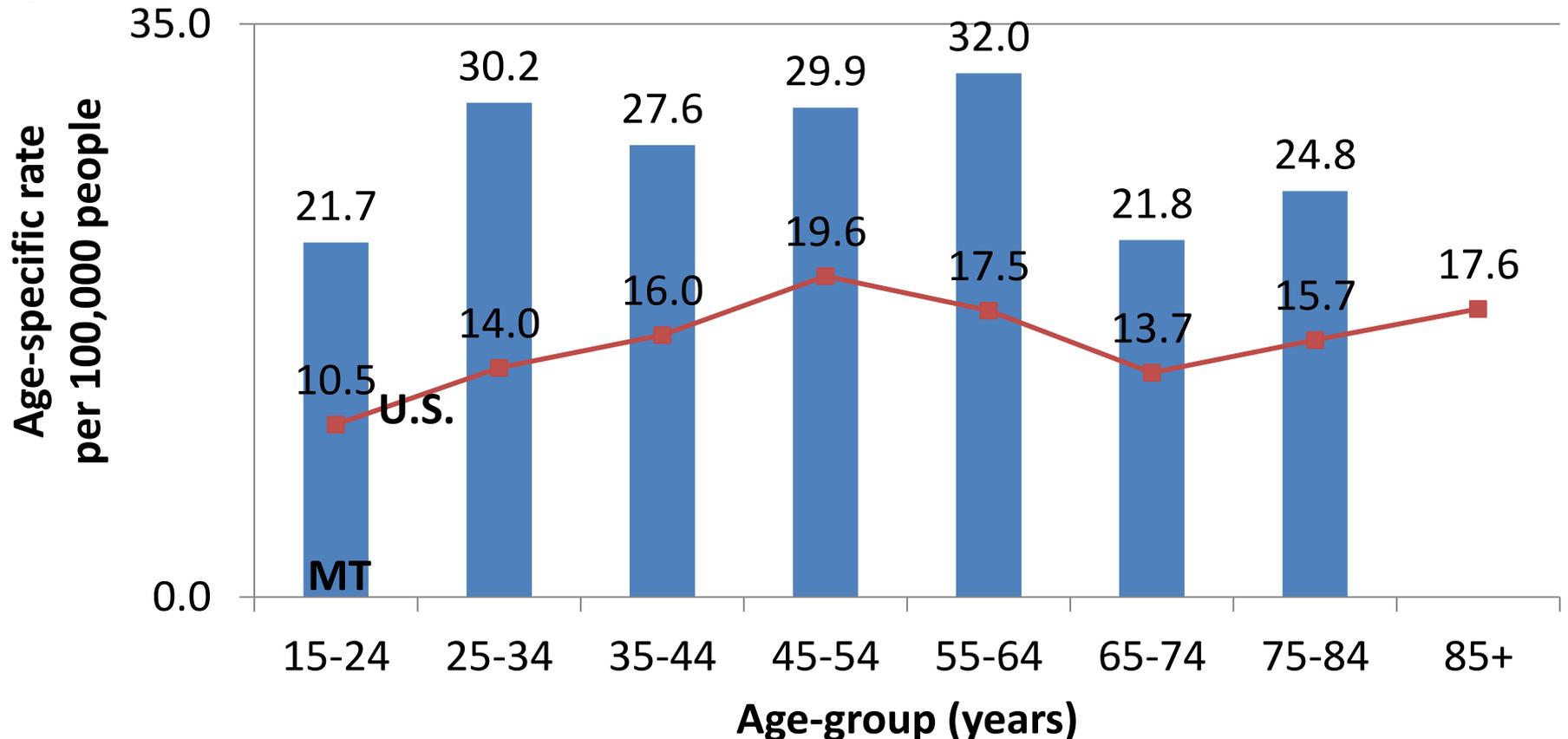
In 2010, Montana suicide rates among youth (15-24) were two times higher than U.S.



TEXT: Labels used sparingly

Suicide was higher in Montana compared to the U.S.

In 2010, Montana suicide rates among youth (15-24) were two times higher than U.S.



ARRANGEMENT

Arrangement

Improper arrangement of graph elements can confuse readers at best and mislead viewer at worst. Thoughtful arrangement makes a data visualization easier for a viewer to interpret.

Proportions are accurate

A viewer should be able to take a ruler to measure the length or area of the graph and find that it matches the relationship in the underlying data.

Data are intentionally ordered

Data should be displayed in an order that makes logical sense to the viewer. Data may be ordered by frequency counts (e.g., from greatest to least for nominal categories), by groupings or bins (e.g., histograms), by time period (e.g., line charts), alphabetically, etc.

Axis intervals are equidistant

The spaces between axis intervals should be the same unit, even if every axis interval isn't labeled.

Graph is two-dimensional

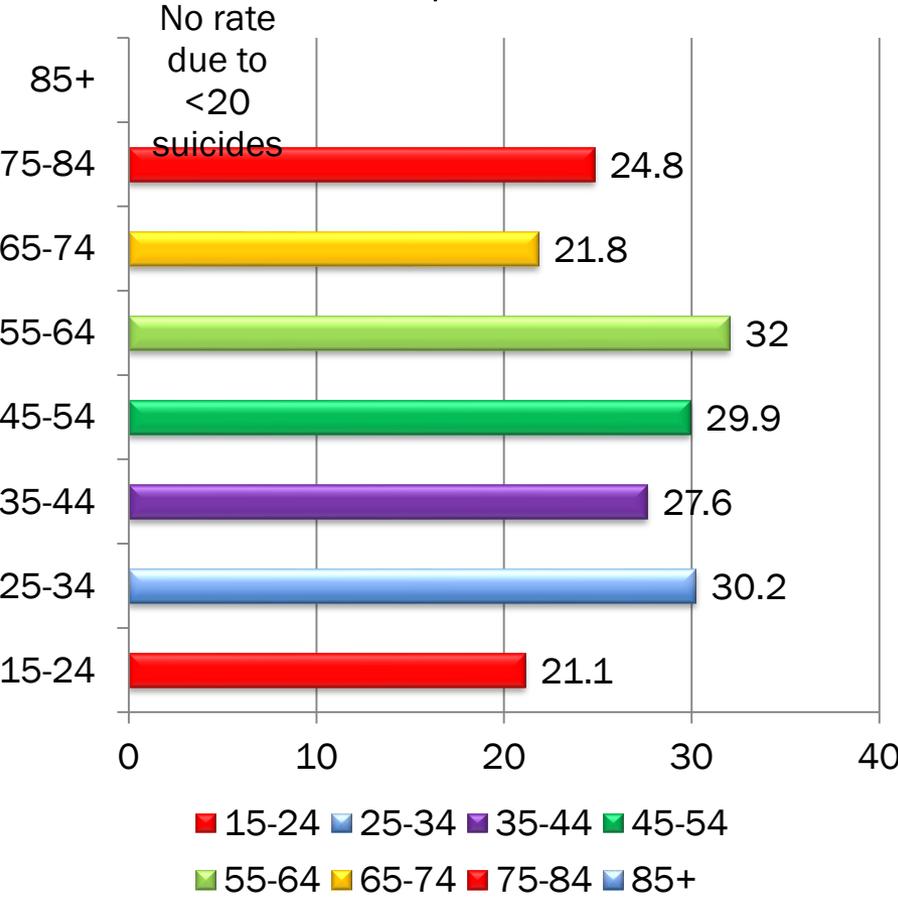
Avoid three-dimensional displays, bevels, and other distortions.

Display is free from decoration

Graph is free from clipart or other illustrations used solely for decoration. Some graphics, like icons, can support interpretation.

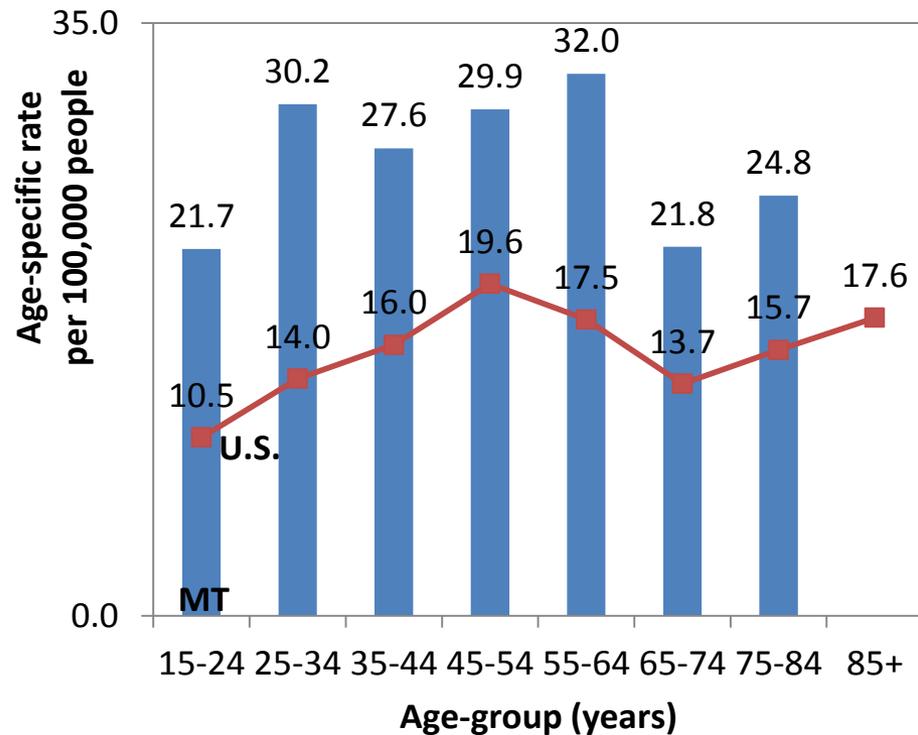
ARRANGEMENT: Data intentionally ordered

Crude suicide Rate by Age, Montana Residents, 2010-2011



Suicide was higher in Montana compared to the U.S.

In 2010, Montana suicide rates among youth (15-24) were two times higher than U.S.



COLOR

Color

Keep culture-laden color connotations in mind. For example, pink is highly associated with feminine qualities in the USA.

Use sites like Color Brewer to find color schemes suitable for reprinting in black-and-white and for colorblindness.

Color scheme is intentional

Colors should represent brand or other intentional choice, not default color schemes. A safe bet for consultants is to use your client's colors. Use online tools to identify brand colors and others that are compatible.

Color is used to highlight key patterns

Action colors should guide the viewer to key parts of the display. Less important or supporting data should be a muted color.

Color is legible when printed in black and white

When printed or photocopied in black and white, the viewer should still be able to see patterns in the data.

Color is legible for people with colorblindness

Avoid red-green and yellow-blue combinations when those colors touch one another.

Text sufficiently contrasts background

Black/very dark text against a white/transparent background is easiest to read.

COLOR: Consider contrast, print quality, and colorblindness

The screenshot displays the ColorBrewer 2.0 web application interface. The main title is "COLORBREWER 2.0" with the subtitle "color advice for cartography". The interface is divided into several sections:

- Number of data classes:** Set to 3.
- Nature of your data:** Radio buttons for "sequential" (selected), "diverging", and "qualitative".
- Pick a color scheme:** Two columns of color swatches labeled "Multi-hue" and "Single hue".
- Only show:** Checkboxes for "colorblind safe", "print friendly", and "photocopy safe".
- Context:** Checkboxes for "roads", "cities", and "borders" (checked).
- Background:** Radio buttons for "solid color" (selected) and "terrain".
- Color transparency:** A slider control.
- 3-class BuGn:** A legend showing three color swatches with their corresponding HEX codes: #e5f5f9, #99d8c9, and #2ca25f.
- EXPORT:** A button for exporting the map.

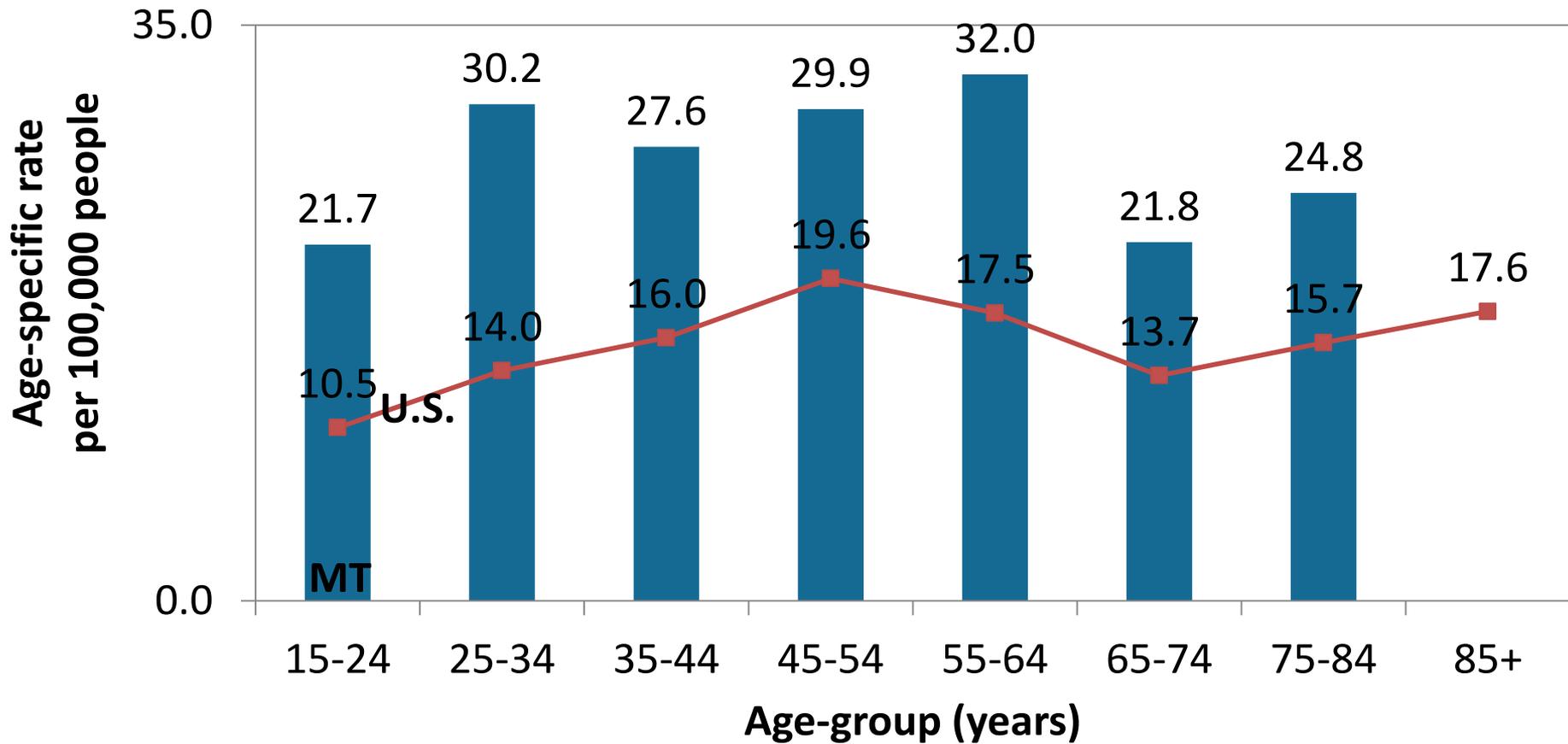
The central map shows a geographical area (Montana) with a 3-class BuGn color scheme applied to its counties. The colors range from light cyan to dark green. The interface also includes a footer with copyright information: "© Cynthia Brewer, Mark Harrower and The Pennsylvania State University Support Back to Flash version Back to ColorBrewer 1.0" and the "axismaps" logo.

colorbrewer2.org

COLOR: Scheme is intentional

Suicide was higher in Montana compared to the U.S.

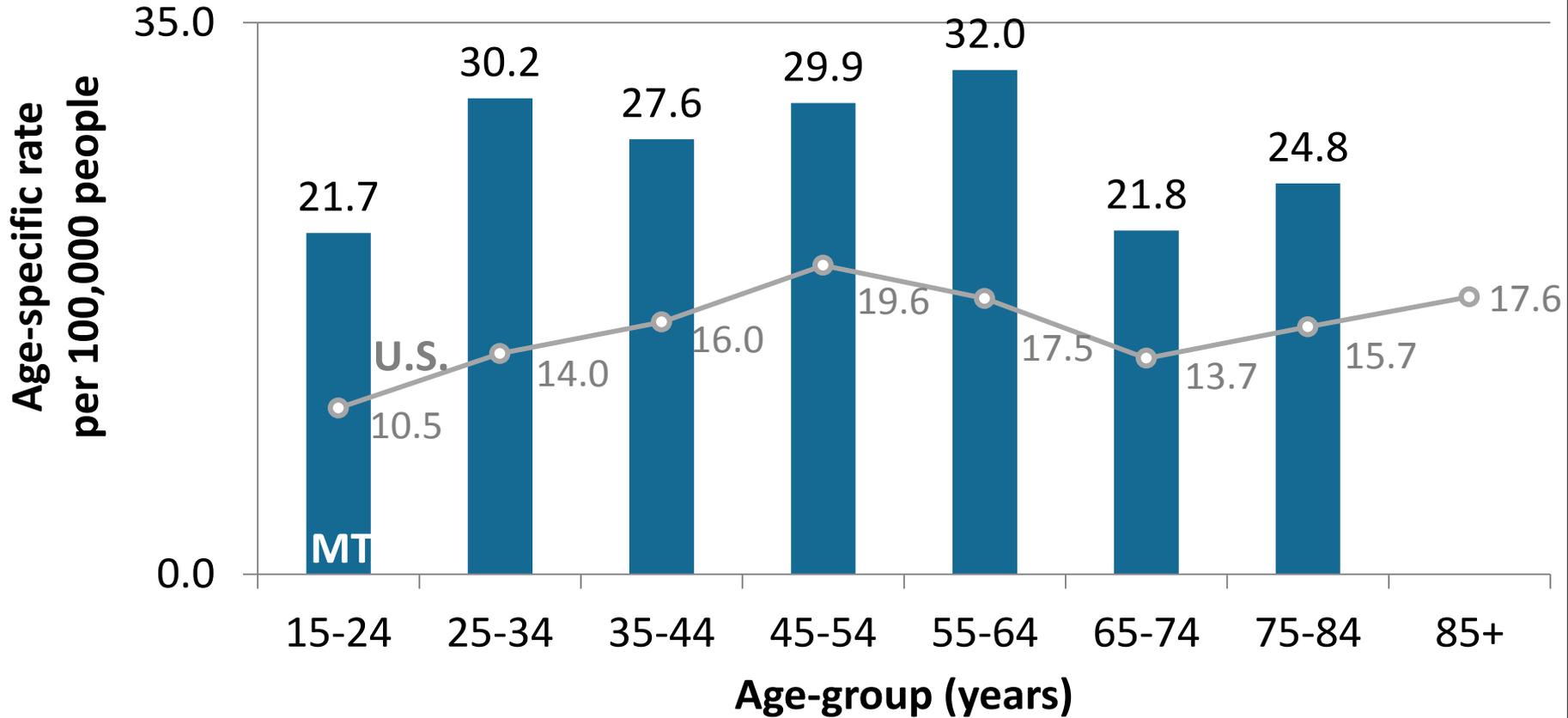
In 2010, Montana suicide rates among youth (15-24) were two times higher than U.S.



COLOR: Use color to highlight key patterns or messages

Suicide was higher in Montana compared to the U.S.

In 2010, Montana suicide rates among youth (15-24) were two times higher than U.S.



LINES

Lines

Excessive lines—gridlines, borders, tick marks, and axes—can add clutter or noise to a graph, so eliminate them whenever they aren't useful for interpreting the data.

Gridlines, if present, are muted

Color should be faint gray, not black. Full points if no gridlines are used.

Graph does not have border line

Graph should bleed into the surrounding page or slide rather than being contained by a border.

Axes do not have unnecessary tick marks

Tick marks are useful in line graphs (to demarcate each point in time along the y-axis) but unnecessary in bar charts.

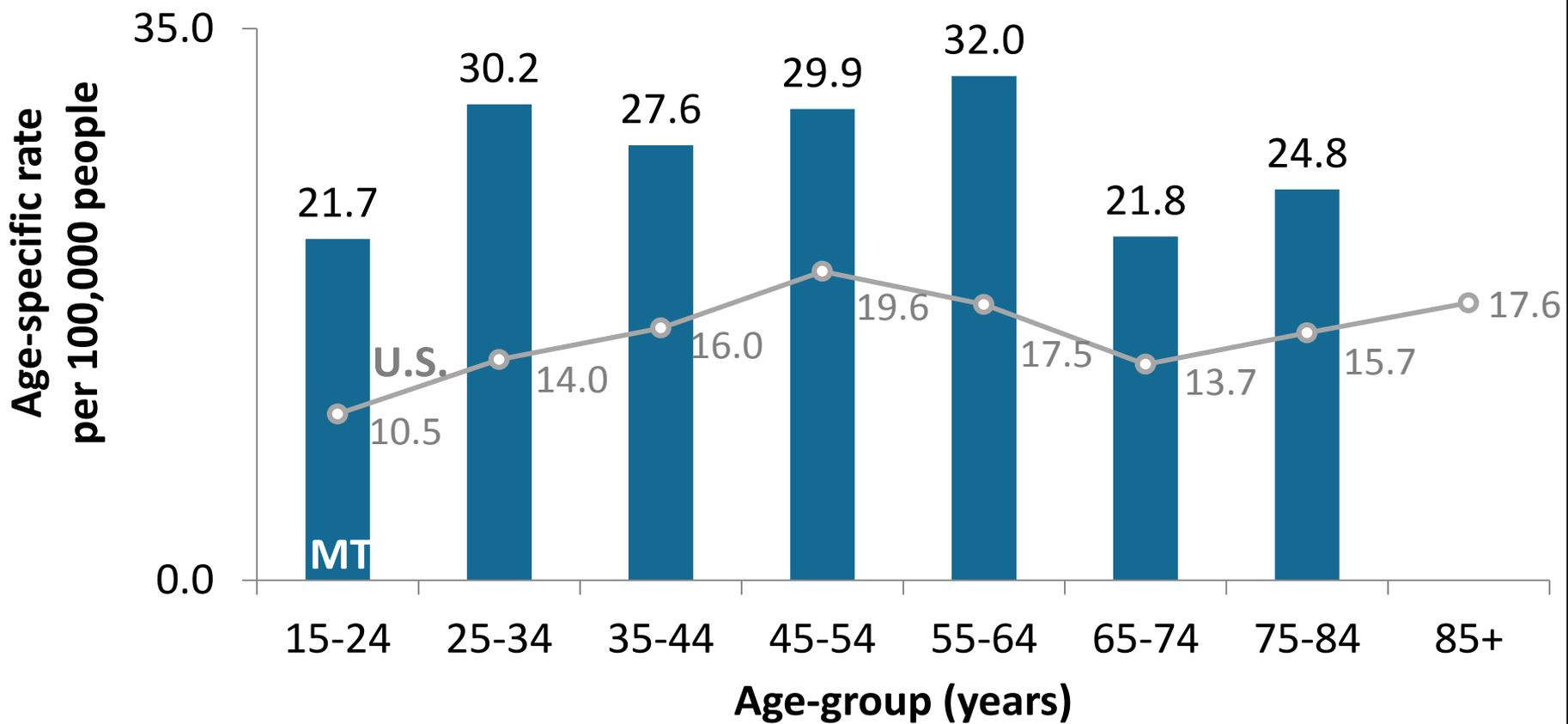
Graph has one horizontal and one vertical axis

Viewers can best interpret one x- and one y-axis, even if one is hidden. Don't add a second y-axis.

LINES: Gridlines, if present, are muted

Suicide was higher in Montana compared to the U.S.

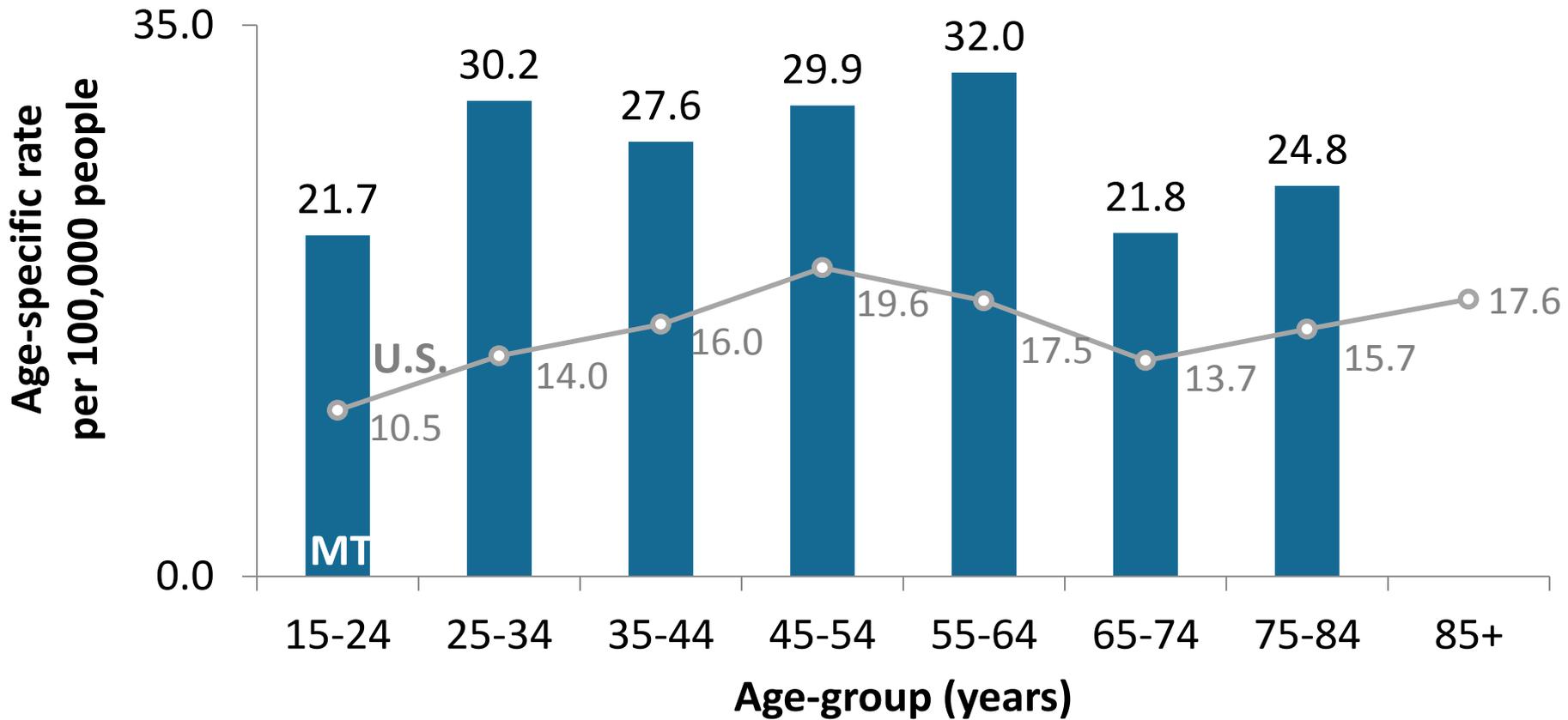
In 2010, Montana suicide rates among youth (15-24) were two times higher than U.S.



LINES: No border around graph

Suicide was higher in Montana compared to the U.S.

In 2010, Montana suicide rates among youth (15-24) were two times higher than U.S.

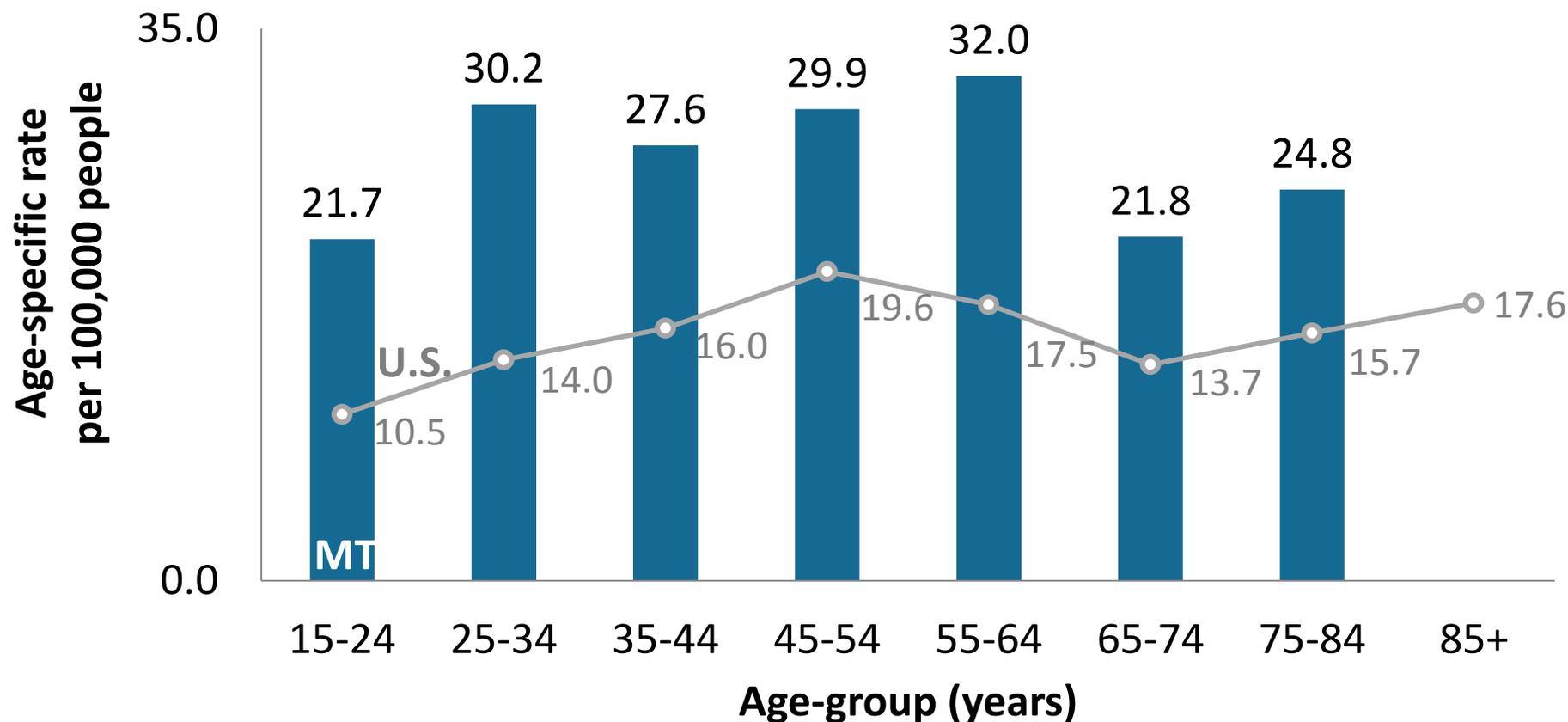


LINES: Remove tick marks from axes

from axes

Suicide was higher in Montana compared to the U.S.

In 2010, Montana suicide rates among youth (15-24) were two times higher than U.S.



OVERALL

Overall

Graphs will catch a viewer's attention so only visualize the data that needs attention. Too many graphics of unimportant information dilute the power of visualization.

Graph highlights significant finding or conclusion

Graphs should have a "so what?" – either a practical or statistical significance (or both) to warrant their presence.

The type of graph is appropriate for data

Data are displayed using a graph type appropriate for the relationship within the data. For example, change over time is displayed as a line graph, area chart, slope graph, or dot plot.

Graph has appropriate level of precision

Few numeric labels need decimal places. When precision is important, choose a type of graph type that displays differences through length or points along a line (e.g., bar charts, dot plots). When precision is less important, you can use a graph that displays differences through angles or area (e.g., pie charts, circle charts).

Contextualized or comparison data are present

Comparisons—over time, across programs or subgroups of participants, etc.—help the viewer understand the significance of the data.

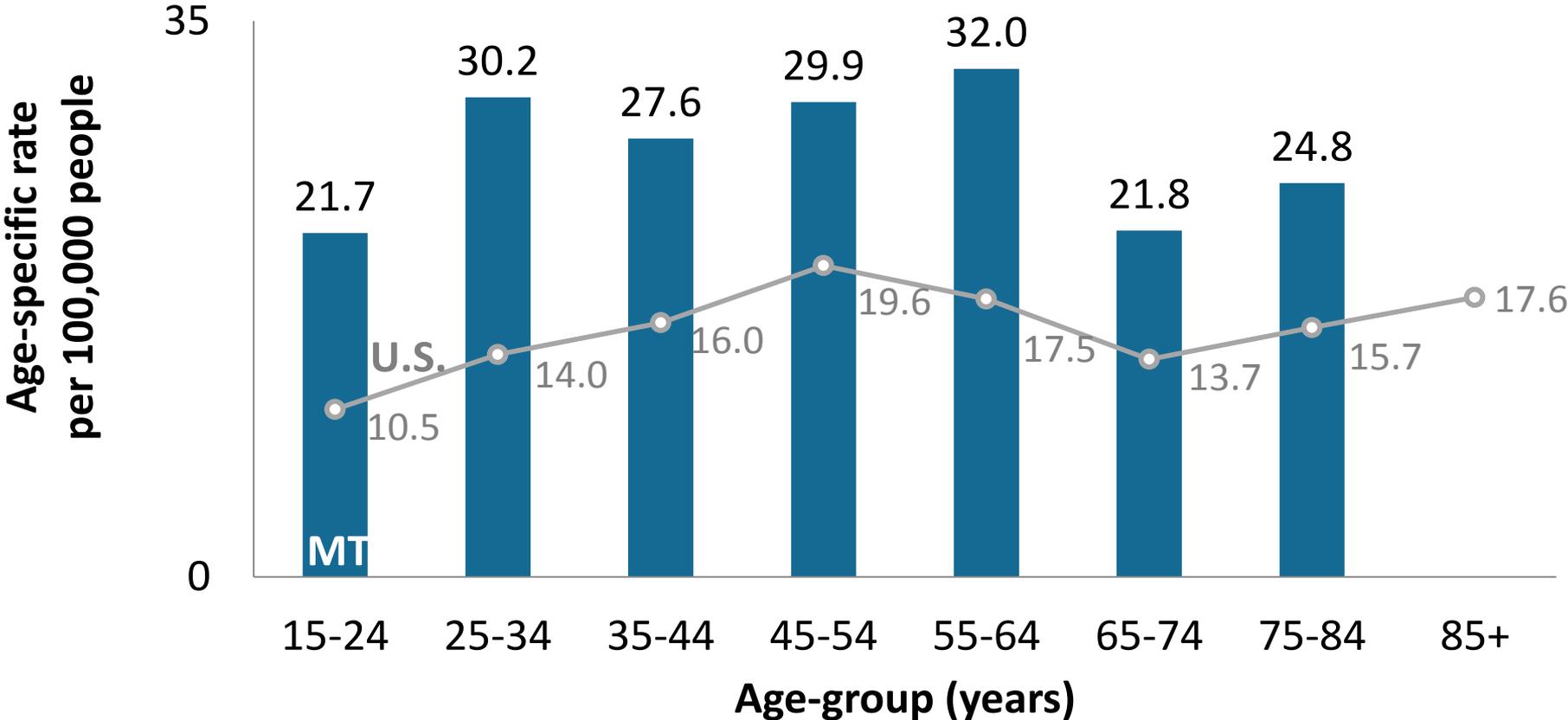
Individual chart elements work together to reinforce the overarching takeaway message

Choices about graph type, text, arrangement, color, and lines should reinforce the same takeaway message.

OVERALL: Precision

Suicide was higher in Montana compared to the U.S.

In 2010, Montana suicide rates among youth (15-24) were two times higher than U.S.

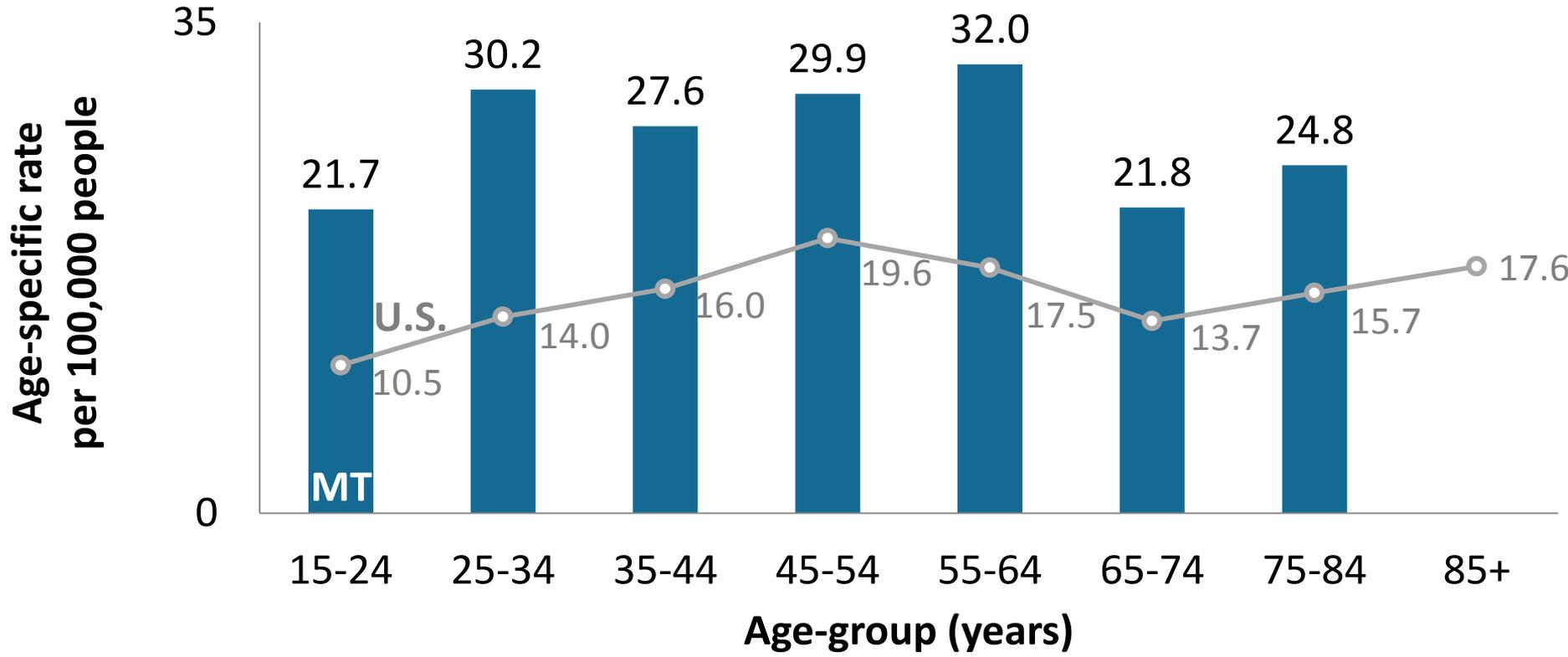


NOT ON CHECKLIST:

Footnote details about data source

Suicide was higher in Montana compared to the U.S.

In 2010, Montana suicide rates among youth (15-24) were two times higher than U.S.



Data source: Montana Death Records, 2010; National Center for Health Statistics, 2010. ←

Footnote

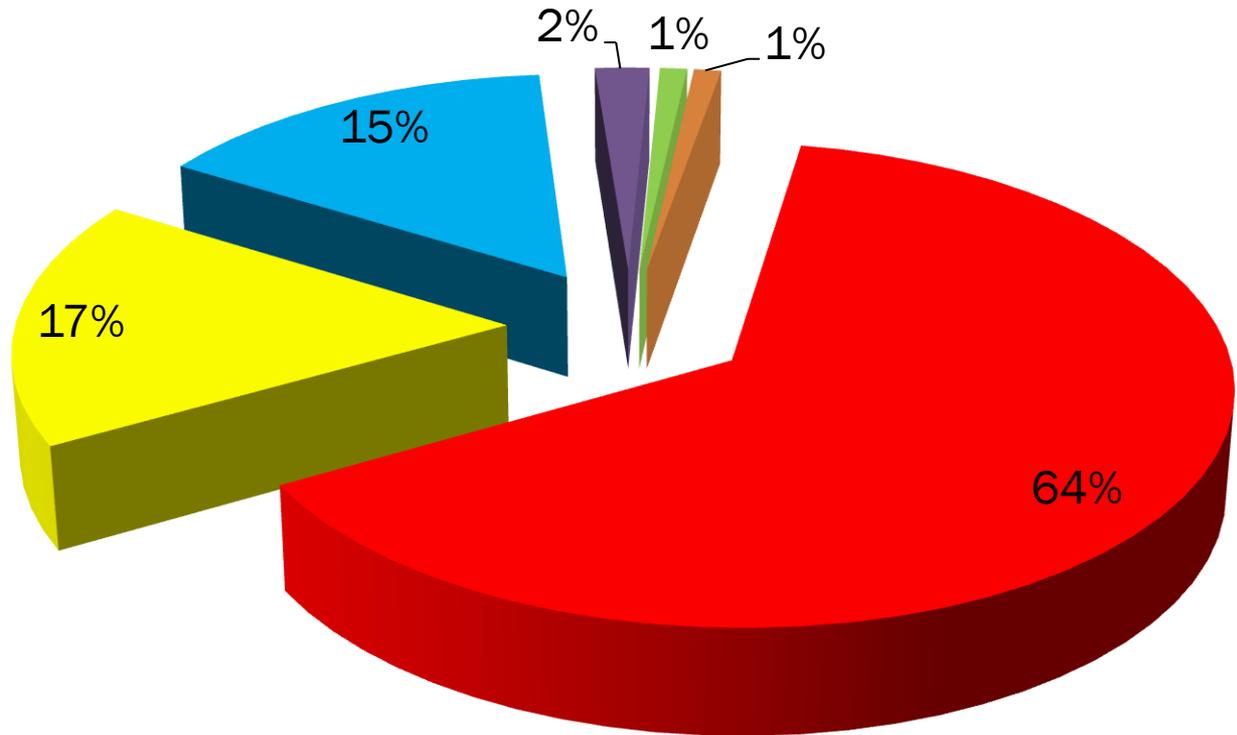
GROUP EXERCISE

FIXING BAD FIGURES

EXAMPLE 1

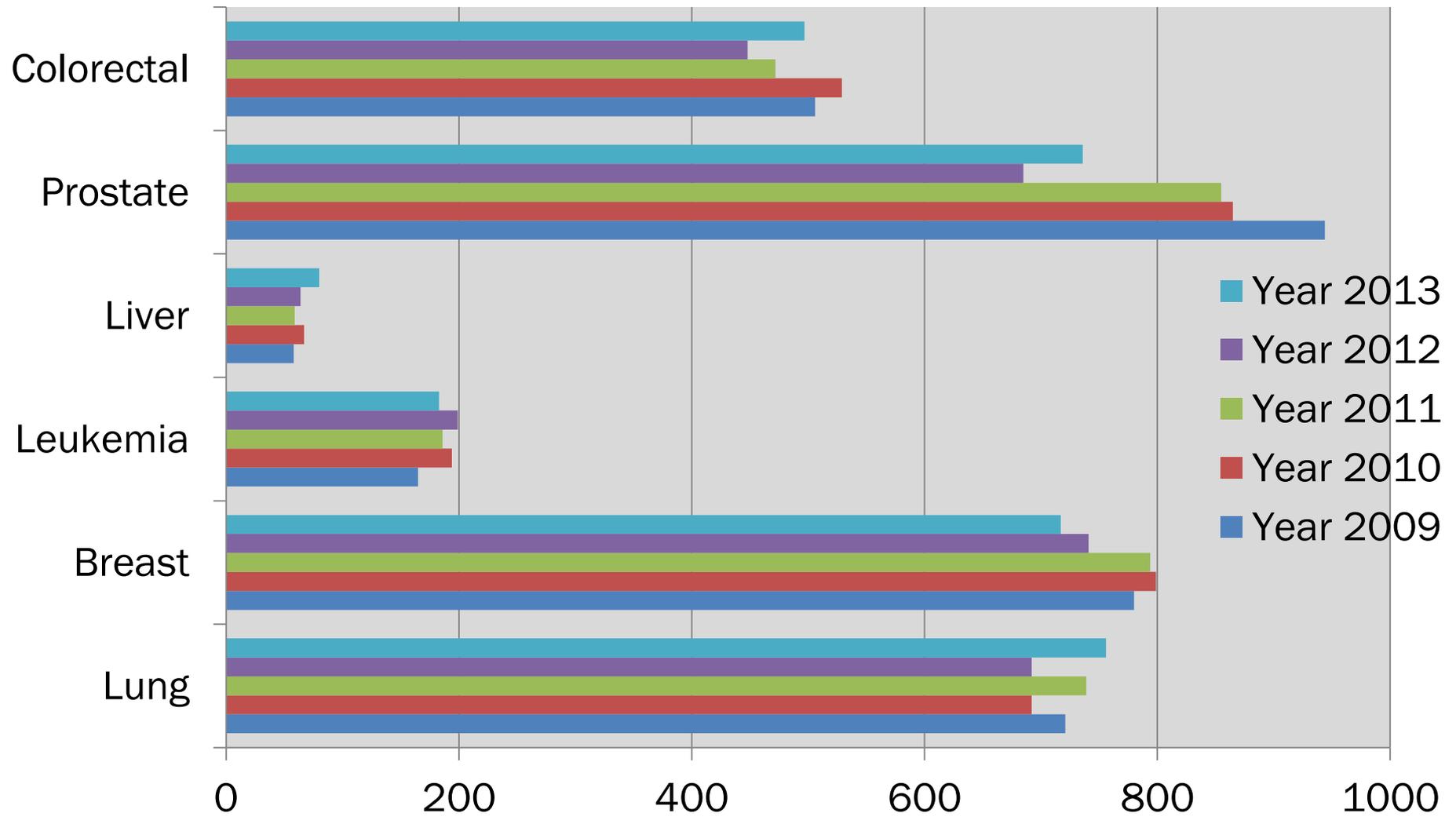
Percent of Suicides by Means, Montana Residents, 2010-2011

■ Firearm ■ Suffocation ■ Poisoning ■ Other ■ Cut or Pierce ■ Drown



EXAMPLE 2

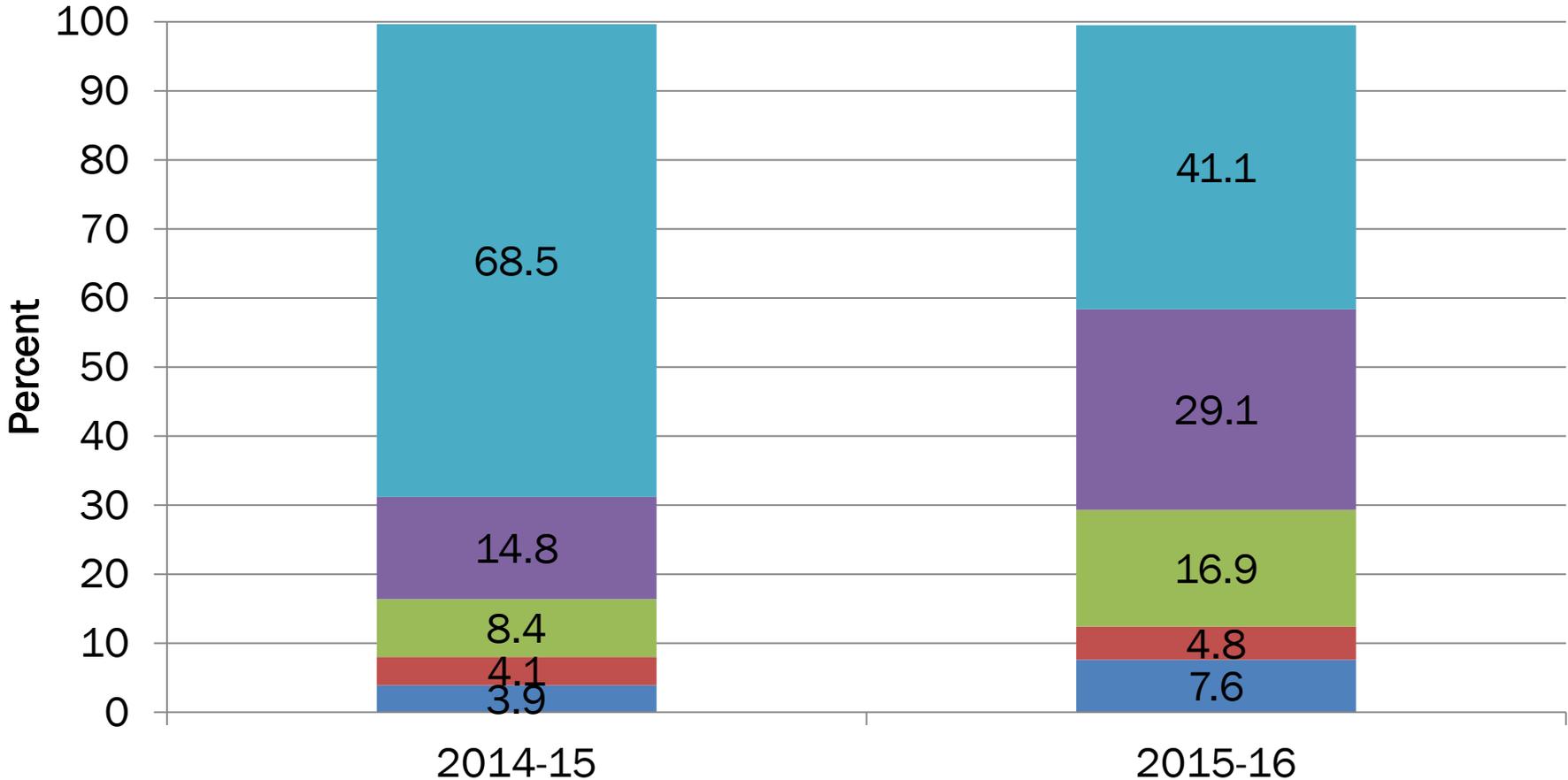
Cancer cases among Montana residents



EXAMPLE 3

Influenza hospitalizations by age group – Montana, 2014-2015 vs. 2015-2016 seasons

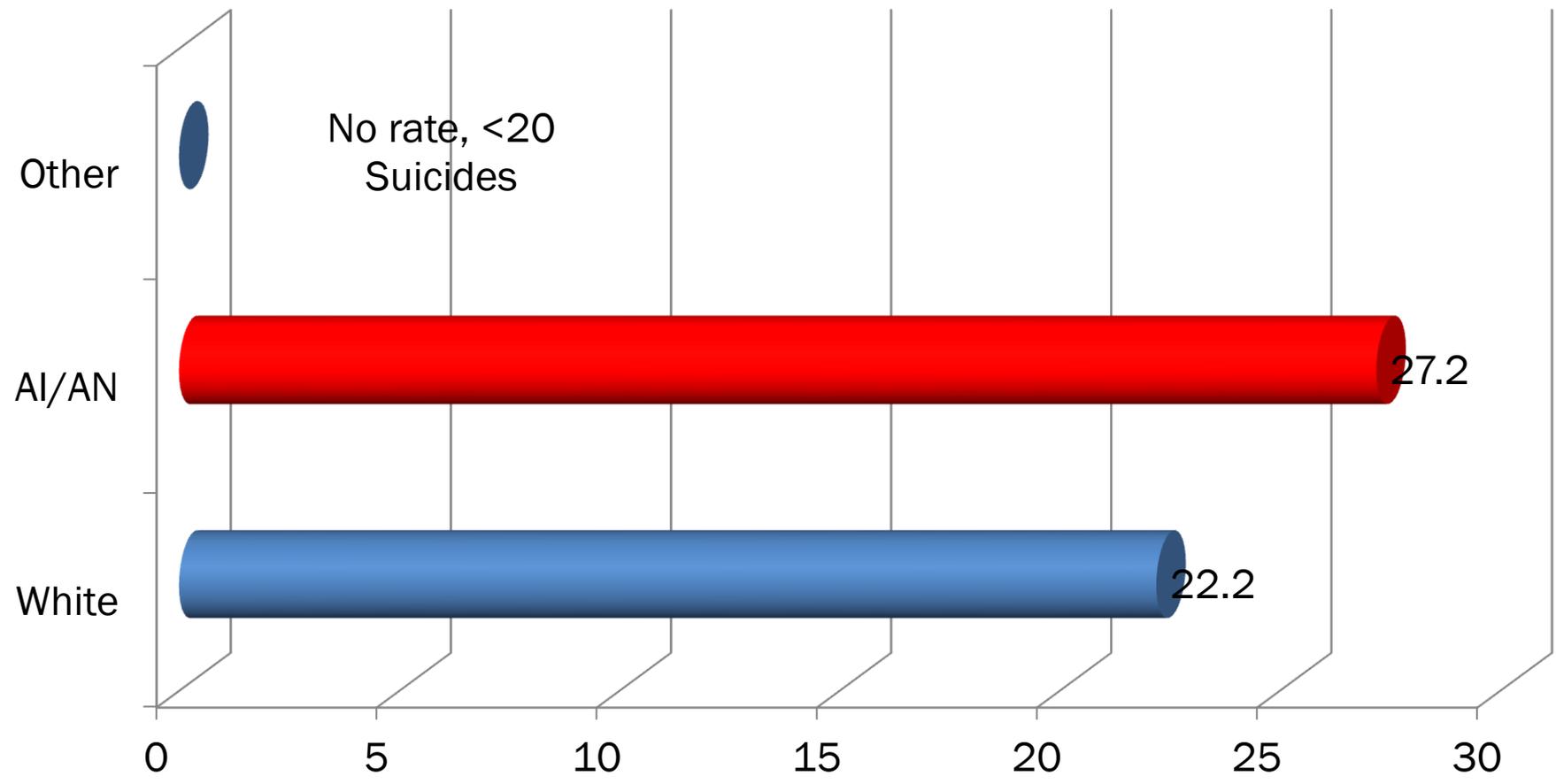
■ 0-4 yr ■ 5-17 yr ■ 18-49 yr ■ 50-64 yr ■ 65+ yr



EXAMPLE 4

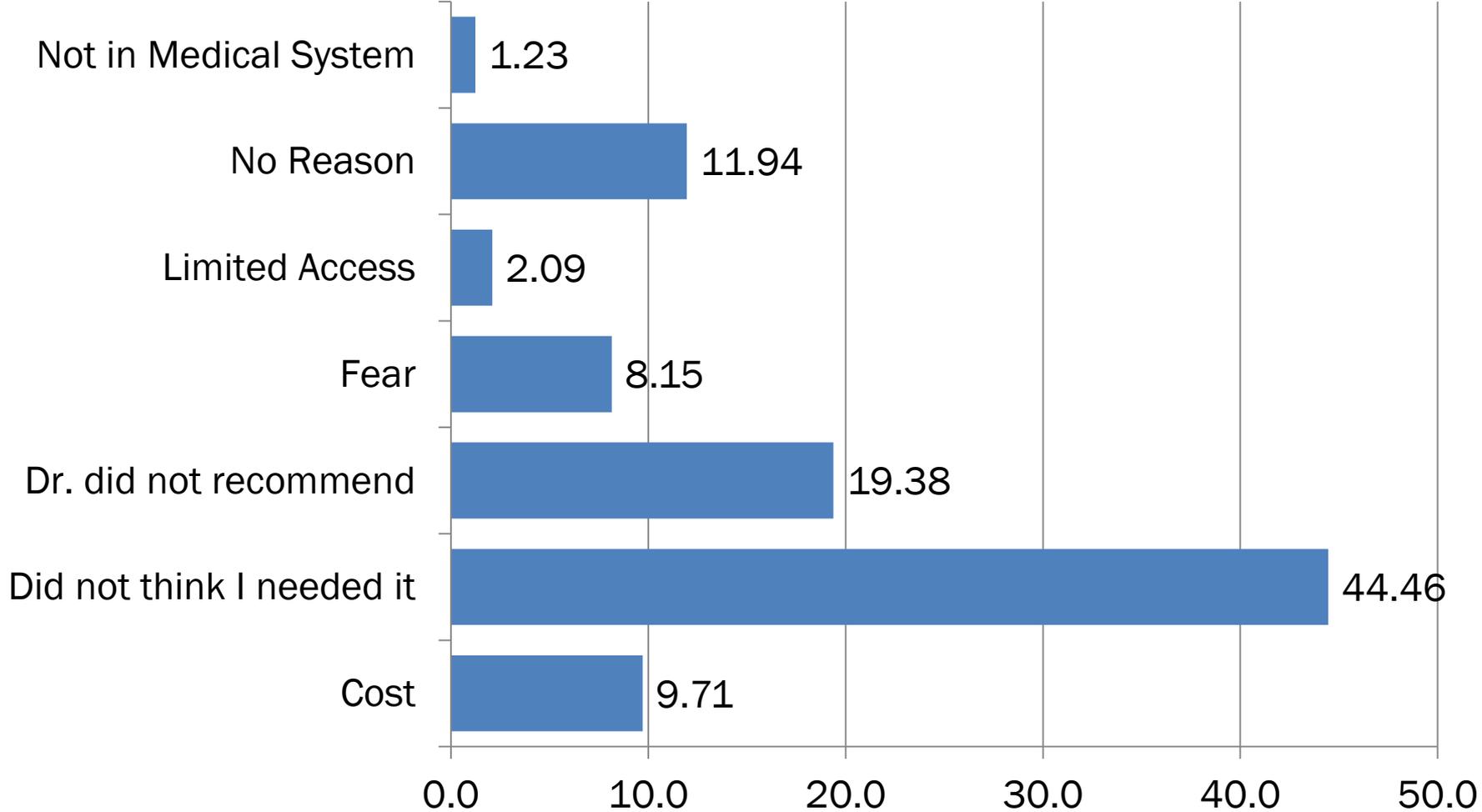
Crude Suicide Rates by Ethnicity, Montana Residents, 2010-2011

■ White ■ AI/AN ■ Other

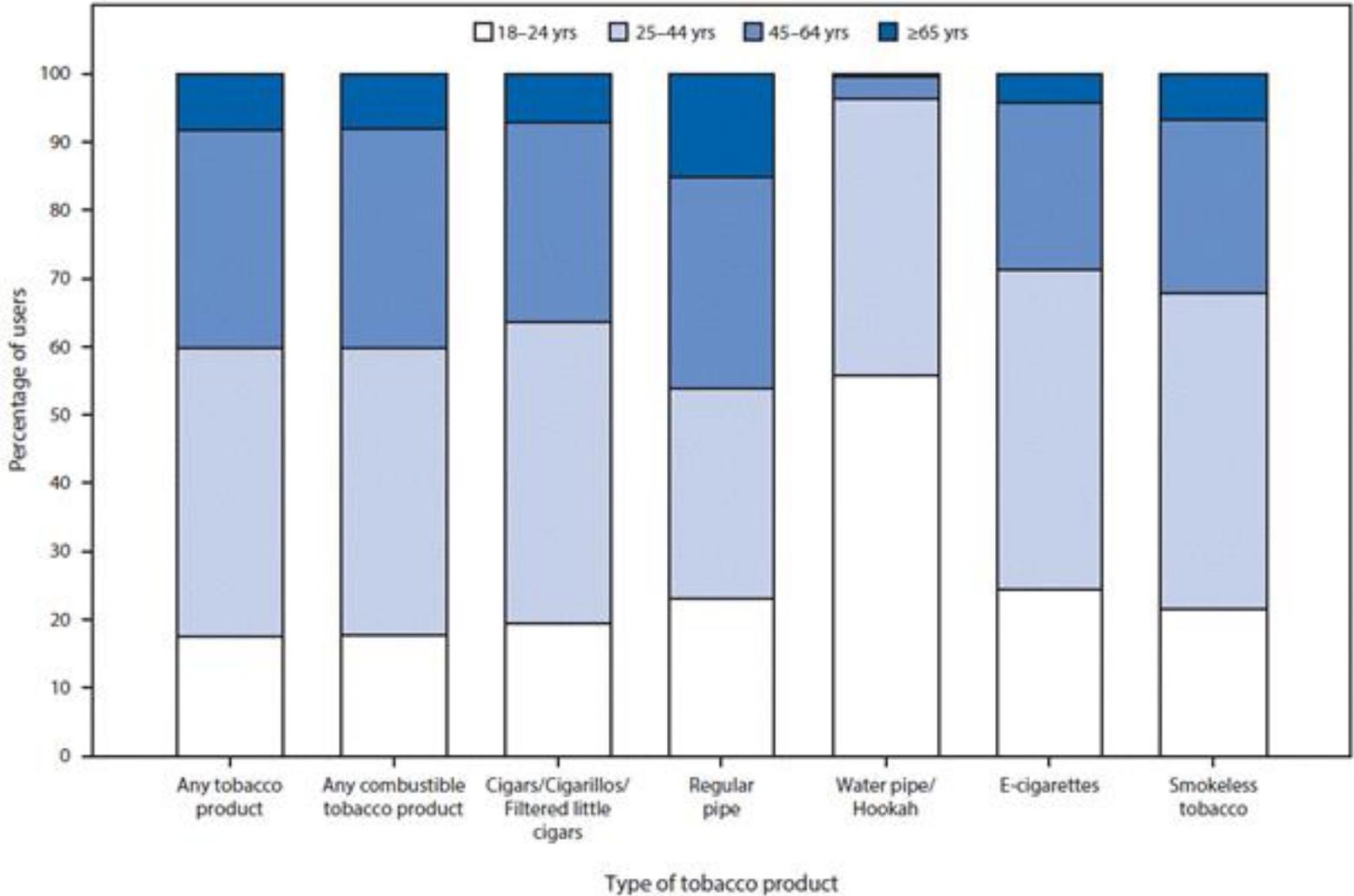


EXAMPLE 5

Main reason Montanas have not had a colonoscopy, 2014



EXAMPLE 6



GOOD DATA VISUALIZATION



Key Stakeholder Survey Results

Key stakeholders prioritized health issues in the following order:

Major problem Moderate problem Minor problem No problem

1 Mental Health

A full 7 in 10 key stakeholders taking part in the online survey characterized mental health as a "major problem" in the community.



"We are severely short on mental health professionals. It is very difficult to impossible to access mental health care for many patients in need. The waiting times to get into the existing mental health professionals is quite long."

– Physician

2 Substance Abuse

The greatest share of key stakeholders taking part in the survey characterized substance abuse as a "moderate problem" in the community.



"People with substance abuse issues too often do not wish to change. Once they decide to try to change, it is not easy to figure out how to access help."

– Public Health

3 Nutrition, Physical Activity and Weight

A large share of key stakeholders taking part in the survey characterized nutrition, physical activity, and weight as a "moderate problem" in the community.

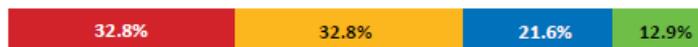


"Our environment is not designed to support individuals making healthy choices. Unhealthy foods are the least expensive. Time constraints and access prevent individuals from getting enough physical activity. It is easier to make unhealthy choices."

– Community/Business Leader

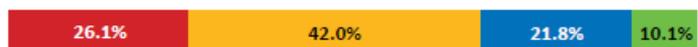
4 Diabetes

Nearly one-third of key stakeholders taking part in the survey characterized diabetes as a "major problem" in the community, and an identical proportion characterized diabetes as a "moderate problem."



5 Heart Disease

The greatest share of key stakeholders taking part in the survey characterized heart disease and stroke as a "moderate problem" in the community.



Other Issues

These issues also were of significant concern to key stakeholders who took the survey:

Tobacco use:
25.8% major problem
46.1% moderate problem

Access to health-care services:
24.8% major problem
50.4% moderate problem

Dementia/Alzheimer's disease:
24.6% major problem
47.5% moderate problem



Percent of Montana adults up-to-date with cancer screenings in 2012 & 2014 and the 2018 goal

Colorectal
men & women
aged 50-75 yrs



Breast
women 50+ yrs



Cervical
women 18+ yrs



50%

100%

Data Source: Montana Behavioral Risk Factor Surveillance System, 2012 & 2014