

## Effective Use of Box Charts

- Purpose** This tool provides guidelines and tips on how to effectively use box charts to communicate research findings.
- Format** This tool provides guidance on box charts and their purposes, and shows examples of preferred practices and practical tips for box charts.
- Audience** This tool is designed primarily for researchers from the Model Systems that are funded by the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR). The tool can be adapted by other NIDILRR-funded grantees and the general public.

The contents of this tool were developed under a grant from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR grant number 90DP0012-01-00). The contents of this fact sheet do not necessarily represent the policy of Department of Health and Human Services, and you should not assume endorsement by the Federal Government.

# Box Charts

- ▶ Box Charts are used to quickly and schematically display the distribution of data, particularly when you wish to compare the data distributions of more than one subgroup.
- ▶ Examples: Display the distribution of weight of adult men versus women, display the distribution of systolic blood pressure for samples of adult men before and after treatment with pressure-lowering medication.
- ▶ Box Charts are also called Box Plots, Box-and-Whisker Charts.

# Box Charts

- ▶ Box Charts are typically oriented vertically, with the numeric values represented across the vertical axis and the categorical groups whose distribution is to be displayed positioned along the horizontal axis.
- ▶ Box Charts have a variety of sub-types, sophistication, and elements.

Box Charts can contain any of the following elements:

- ▶ Numeric value at the Median [50<sup>th</sup> percentile] represented by a center line in the box.
- ▶ A separate symbol within the box might designate the arithmetic mean.
- ▶ Numeric value at the 25<sup>th</sup> and 75<sup>th</sup> percentiles – represented by the closed bottom and top lines of the box, respectively.
- ▶ 25<sup>th</sup> percentile labelled Q1. 75<sup>th</sup> percentile labelled Q3.
- ▶ The 25<sup>th</sup> and 75<sup>th</sup> percentile marks are often called the “hinges” of the box.
- ▶ The Inter-Quartile Range [IQR] is the value represented by the 75<sup>th</sup> percentile minus the value represented by the 25<sup>th</sup> percentile.

# Box Charts

Box Charts can contain any of the following elements (continued):

▶ Range Indicators:

- Whiskers – vertical lines extending from the box hinge markers toward values at the high and low end of the distribution, beyond which any data points might be considered “outliers”.
- The values at the whiskers high and low end boundaries are typically marked with “T” caps on the whisker lines.
- Several approaches to calculating the Whiskers high and low values:
- A common method is to set the whiskers boundaries at 1.5 times the interquartile range  $[Q1 - 1.5 * IQR]$   $[Q3 + 1.5 * IQR]$ .
- One problem with defining the whisker length with  $1.5 * IQR$  is that sometimes the thus-calculated whisker end points are lower than or higher than the actual Minimum and Maximum values in the data series.
- In some Box Charts, the 10<sup>th</sup> percentile and the 90<sup>th</sup> percentile are considered the Whisker end points.
- In some Box Charts, the minimum and maximum values themselves are considered the Whisker end points.

▶ The actual High and Low values of the distribution – typically represented as dots. Extreme high and low values are considered outliers.

# Box Charts

Box Charts can contain any of the following elements (continued):

▶ Error Indicators:

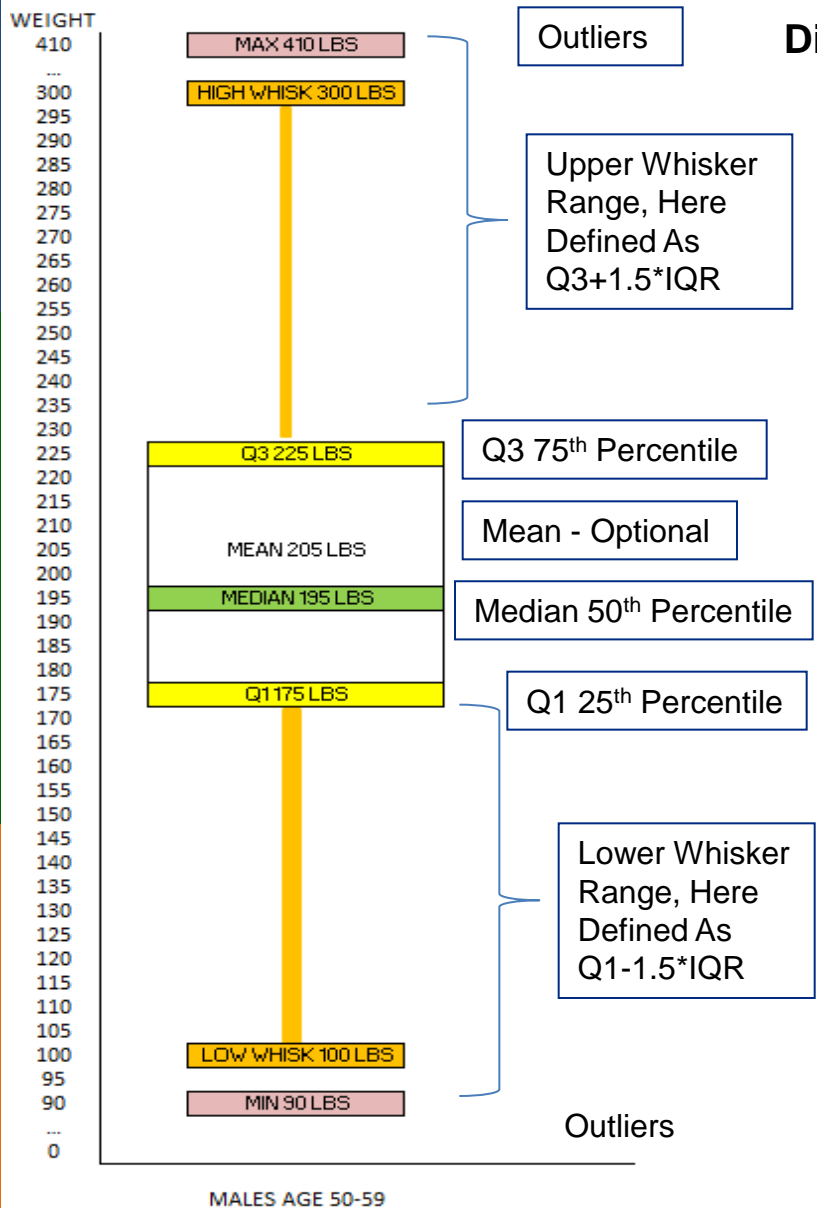
- Other Box Charts may define the whisker ranges in terms of:
- Plus or minus 1 standard deviation of the arithmetic mean
- Plus or minus 2 standard deviations
- Standard error calculations
- 95% and 99% confidence intervals
- Any of which may be represented by whiskers-like capped line boundary indicators.

Because of the variation in the range and error approaches to Box Charts, it is always helpful to precisely define the approach used, in the caption of the Box Chart.

Similarly, look for such interpretative statements when evaluating a Box Chart in a presentation or document.

# Box Charts – Weight Distribution Males Age 50-59 US 2007-2008

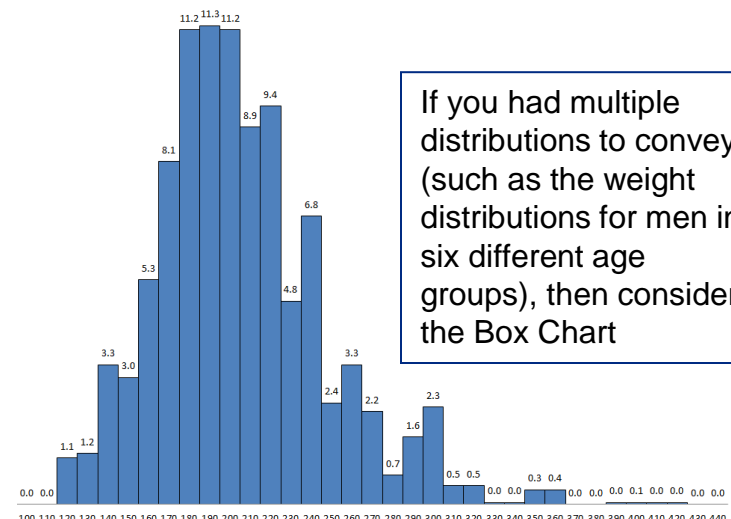
## Stylized Box Chart of Weight Distribution for Males Age 50-59



**Note:** Since the outliers (especially on the high weight end) were so distant from the rest of the chart, the vertical axis was split so as to fit the chart on the page.

Can determine that this distribution is somewhat positively skewed as the mean is pulled toward the higher weight values

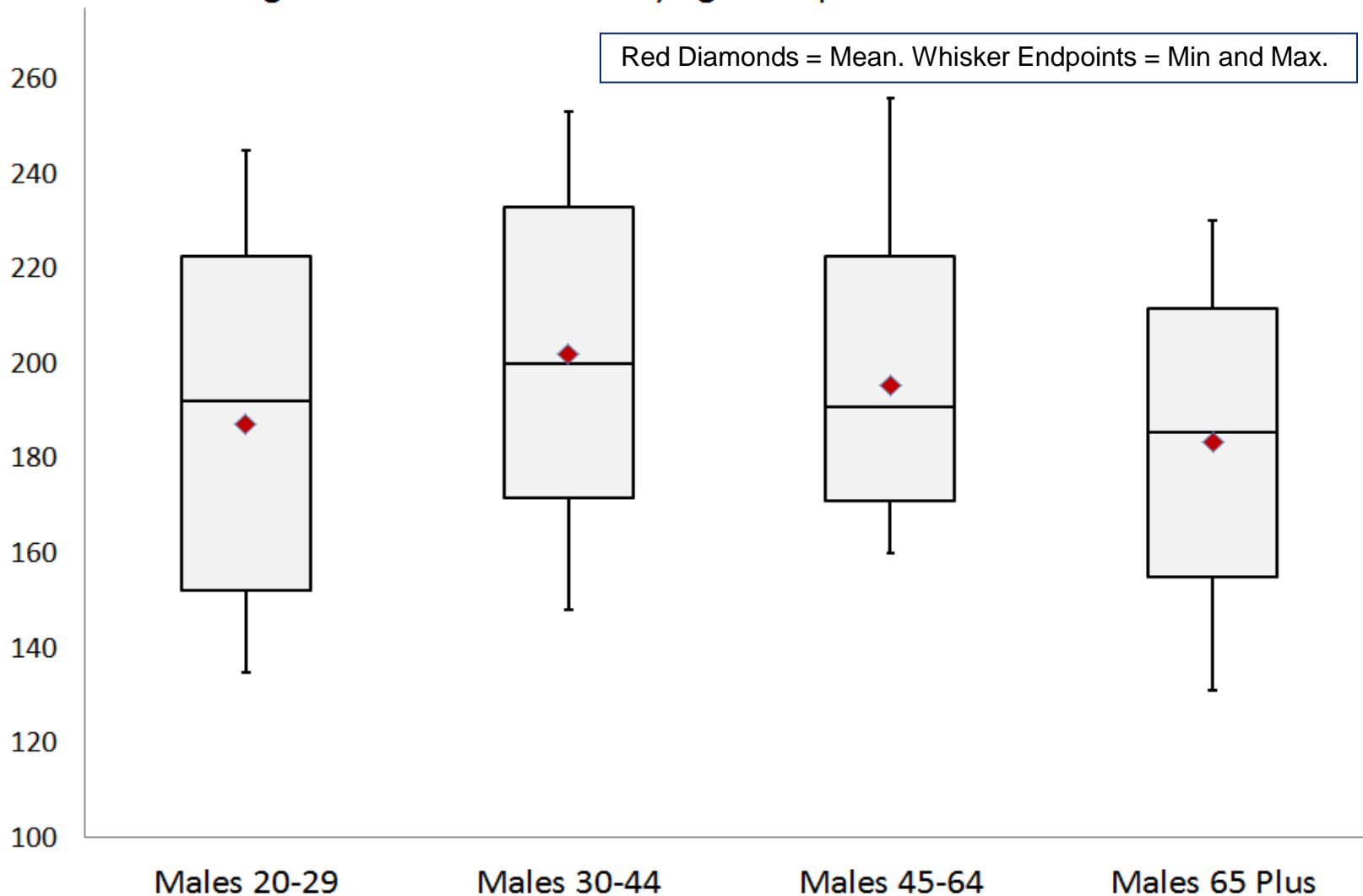
Below is the actual weight histogram for males 50-59. If you had only one distribution to convey, a histogram would be appropriate.



If you had multiple distributions to convey (such as the weight distributions for men in six different age groups), then consider the Box Chart

# Box Charts

Weight Distribution - Males - By Age Group - North Carolina - 2014



# Box Charts

## Systolic Blood Pressure Before and After Treatment Regimen - By Gender

