HRSA Ryan White HIV/AIDS Program
CENTER FOR QUALITY IMPROVEMENT & INNOVATION

CQII Learning Lab

Histograms

HRSA Ryan White HIV/AIDS Program
CENTER FOR QUALITY IMPROVEMENT & INNOVATION

Learning Objectives

After participating in this presentation , you will have a better understanding of:

- What a histogram is and the elements that comprise it
- The uses of the histogram
- How to make a histogram

History of Histograms

The statistician Karl Pearson (1895) first proposed the idea of a historical diagram. The term evolved into *histogram*

What Is a Histogram

- They are similar to bar graphs but they deal with frequency distribution, not raw numbers
  - Bar graphs display raw numbers; sometimes comparing variables (number of males versus females)
  - Histograms display the number of times something occurs using one variable (age: How many clients are between 20 -29; 30 -39, etc.)
- They are used for continuous measures; one variable with a number of different values
  - Ex: age is a measure in which you have multiple values
  - Ex: viral suppression of each patient
- The more data the better; remember it's a frequency distribution

Why Histograms?

- Very useful for larger data sets (50 observations or more)
- Gives a visual picture of the distribution of the data
- Clearly displays any skewing in the data (to the left or right)
- It's a useful decision making tool
- Identifies outliers

Let's Examine Some Terms

- Continuous variable – think of age is a variable. There are a range of values within age but its one variable with multiple values
- Bin – also called a class or an interval. The bin describes its membership
  - Example – one of the age bins could be 18 to 24 year old individuals
  - Think of multiple baskets of apples; one basket has 20 apples, one has 40, etc. The basket is the bin.
- Frequency – This is the number of times a variable occurs within a bin such as 20 apples, having 16 clients between the ages of 128 to 24 years old

How To Construct a Histogram

- First, determine how many bins to use and consider
  - To few will clump the data together
  - To many will spread it out too much
  - There is no "correct" number of bins; usually there are between 3 and 7 bins
- Next, calculate how to distribute the data in each bin; let's use age as an example
  - Collect the age range of each consumer
  - Youngest is 18, oldest is 49
  - Let's use 6 bins for this example (arbitrarily chosen)
  - Find the difference between the oldest and youngest consumer; in this case its 31 years (49-18)

How To Construct a Histogram

- Now divide the 31 by the number of bins you want and you will get $5.1\overline{6}$
- Can you use that number? No, of course not
- Always round UP when you have a fraction; now we have 6
- This 6 represents the range of ages in each bin; not the number of bins

- Let's see how this works based on what we know
  - Our youngest consumer is 18; that's the starting point
  - The age range for our first bin is then 18 to 23 (18,19,20,21,22,23)
  - We then calculate the rest of them until we get to 6 bins

How To Construct a Histogram

- Next, we enter our data which will look like this:

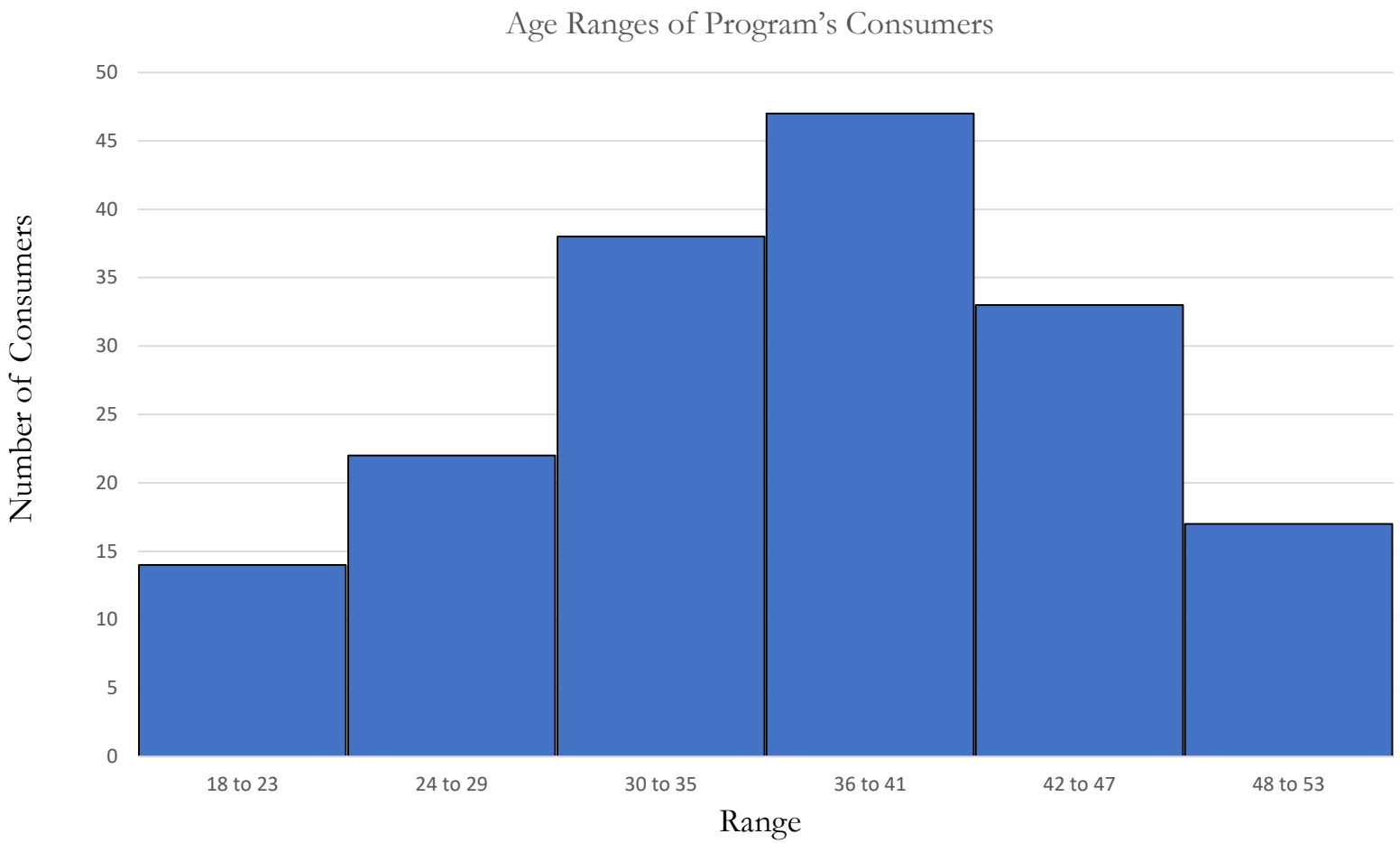
Bin	Frequency
18 to 23	14
24 to 29	22
30 to 35	38
36 to 41	47
42 to 47	33
48 to 53	17

- From here, we can build the Histogram

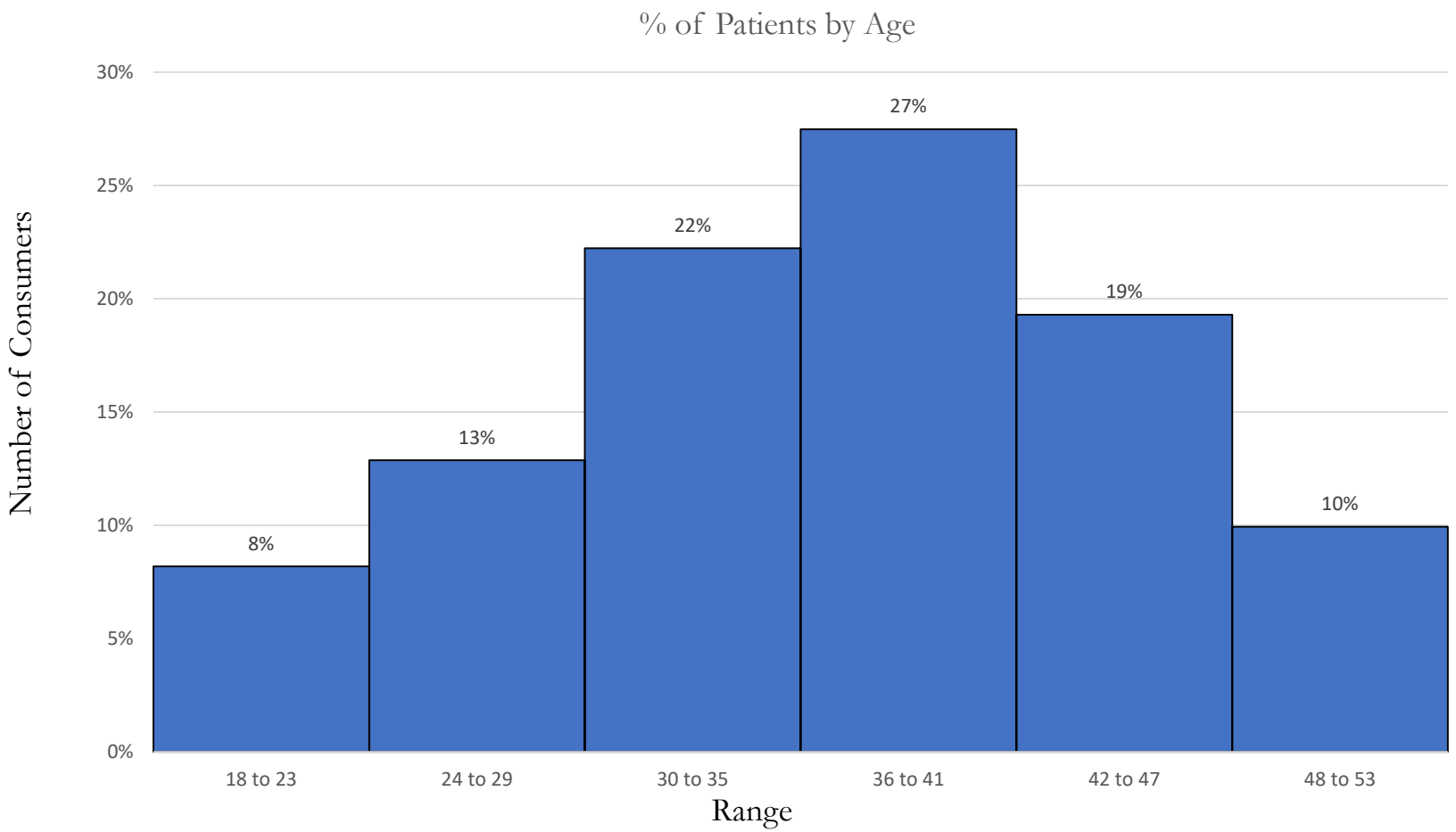
Types of Histograms

- There are two main types of histograms
  - Frequency histogram – the height of a bar represents the frequency or the number of times a measure is represented
  - Relative frequency histogram – the frequency is expressed as a percent of the whole
- Choose the one that will best display what you want to examine

Frequency Histogram



Relative Frequency Histogram



Applications

- Gain a better understanding of your consumer demographics
- Get a better idea of who has higher suppression rates; then find out why
- They can identify outliers in your data
- Better at displaying large data sets into easy to understand graphics
- They will communicate information in a more easy to understand format - graphics

Resources

- ASQ.Org has a histogram template available for download. Go to:

https://asq.org/-/media/public/learn-about-quality/data-collection-analysis-tools/data-point-histogram.xls?la=en

- Southeastern Louisiana University has a great online resource https://www2.southeastern.edu/Academics/Faculty/dgurney/Math241/StatTopics/HistGen.htm
- Count Numbers by range in Excel

https://exceljet.net/formula/count-numbers-by-range-with-countifs

- Statistics How to site

https://www.statisticshowto.datasciencecentral.com/choose-bin-sizes-statistics/