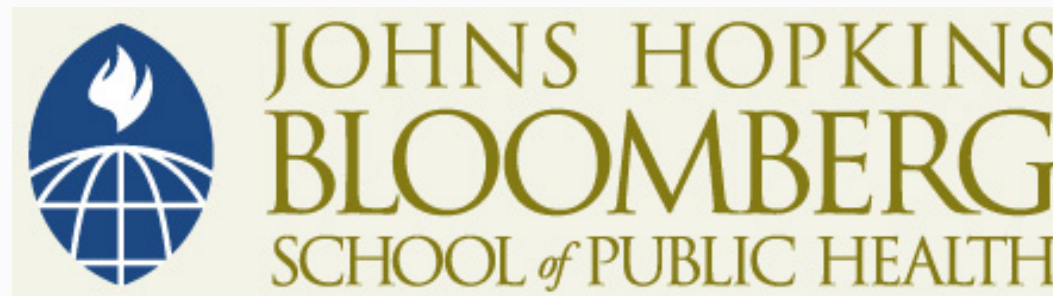


This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike License](https://creativecommons.org/licenses/by-nc-sa/4.0/). Your use of this material constitutes acceptance of that license and the conditions of use of materials on this site.



Copyright 2009, The Johns Hopkins University and John McGready. All rights reserved. Use of these materials permitted only in accordance with license rights granted. Materials provided "AS IS"; no representations or warranties provided. User assumes all responsibility for use, and all liability related thereto, and must independently review all materials for accuracy and efficacy. May contain materials owned by others. User is responsible for obtaining permissions for use from third parties as needed.



JOHNS HOPKINS  
BLOOMBERG  
SCHOOL *of* PUBLIC HEALTH

## Section D

---

### Visually Displaying Continuous Data: Histograms

# Pictures of Data: Continuous Variables

- Histograms
  - Means and medians and standard deviations do not tell the whole story
  - Differences in shape of the distribution
  - Histograms are a way of displaying the distribution of a set of data by charting the number (or percentage) of observations whose values fall within pre-defined numerical ranges

# How to Make a Histogram

- Consider the following data collected from the 1995 Statistical Abstracts of the United States
  - For each of the 50 United States, the proportion of individuals over 65 years of age has been recorded

# How to Make a Histogram

State	%	State	%	State	%	State	%
AL	12.9	IN	12.8	NE	14.1	SC	11.9
AK	4.6	IA	15.4	NV	11.3	SD	14.7
AZ	13.4	KS	13.9	NH	11.9	TN	12.7
AR	14.8	KY	12.7	NJ	13.2	TX	10.2
CA	10.6	LA	11.4	NM	11.0	UT	8.8
CO	10.1	ME	13.9	NY	13.2	VT	12.1
CT	14.2	MD	11.2	NC	12.5	VA	11.1
DE	12.7	MA	14.1	ND	14.7	WA	11.6
FL	18.4	MI	12.4	OH	13.4	WV	15.4
GA	10.1	MN	12.5	OK	13.6	WI	13.4
HI	12.1	MI	12.5	OR	13.7	WY	11.1
ID	11.6	MO	14.1	PA	15.9		
IL	12.6	MT	13.3	RI	15.6		

Source: Statistical Abstracts of the United States (1995).

# How to Make a Histogram

State	%	State	%	State	%	State	%
AL	12.9	IN	12.8	NE	14.1	SC	11.9
AK	4.6	IA	15.4	NV	11.3	SD	14.7
AZ	13.4	KS	13.9	NH	11.9	TN	12.7
AR	14.8	KY	12.7	NJ	13.2	TX	10.2
CA	10.6	LA	11.4	NM	11.0	UT	8.8
CO	10.1	ME	13.9	NY	13.2	VT	12.1
CT	14.2	MD	11.2	NC	12.5	VA	11.1
DE	12.7	MA	14.1	ND	14.7	WA	11.6
FL	18.4	MI	12.4	OH	13.4	WV	15.4
GA	10.1	MN	12.5	OK	13.6	WI	13.4
HI	12.1	MI	12.5	OR	13.7	WY	11.1
ID	11.6	MO	14.1	PA	15.9		
IL	12.6	MT	13.3	RI	15.6		

Source: Statistical Abstracts of the United States (1995).

## How to Make a Histogram

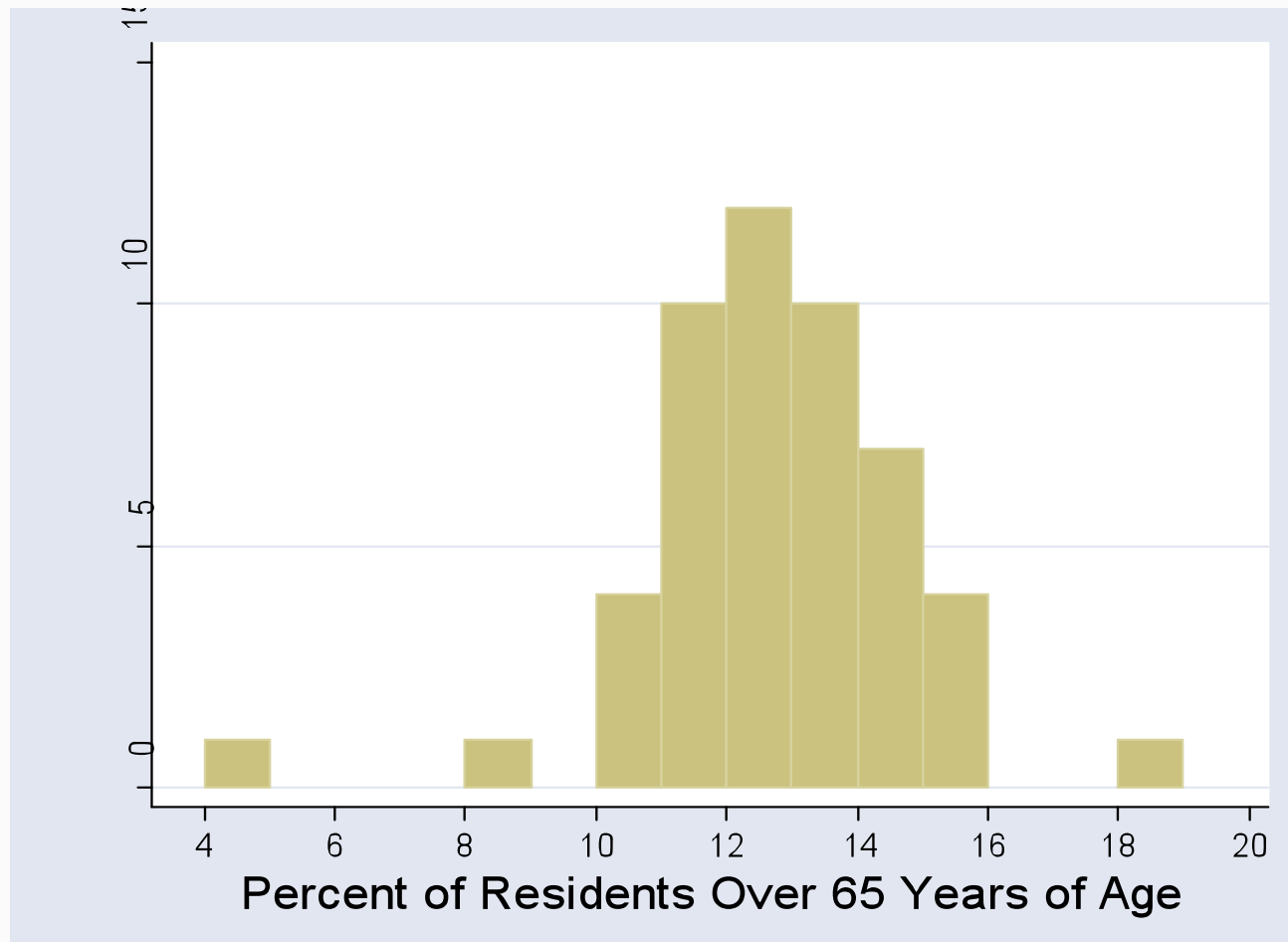
- Break the data range into mutually exclusive, equally sized “bins”:  
here each is 1% wide
- Count the number of observations in each bin

Class	Count	Class	Count	Class	Count
[4.0-5.0]	1	[9.0-10.0]	0	[14.0-15.0]	7
[5.0-6.0]	0	[10.0-11.0]	5	[15.0-16.0]	4
[6.0-7.0]	0	[11.0-12.0]	9	[16.0-17.0]	0
[7.0-8.0]	0	[12.0-13.0]	12	[17.0-18.0]	0
[8.0-9.0]	1	[13.0-14.0]	10	[18.0-19.0]	1

$$\bar{x} = 12.7\%; s = 2.1\%$$

# How to Make a Histogram

- Draw the histogram
- Label scales

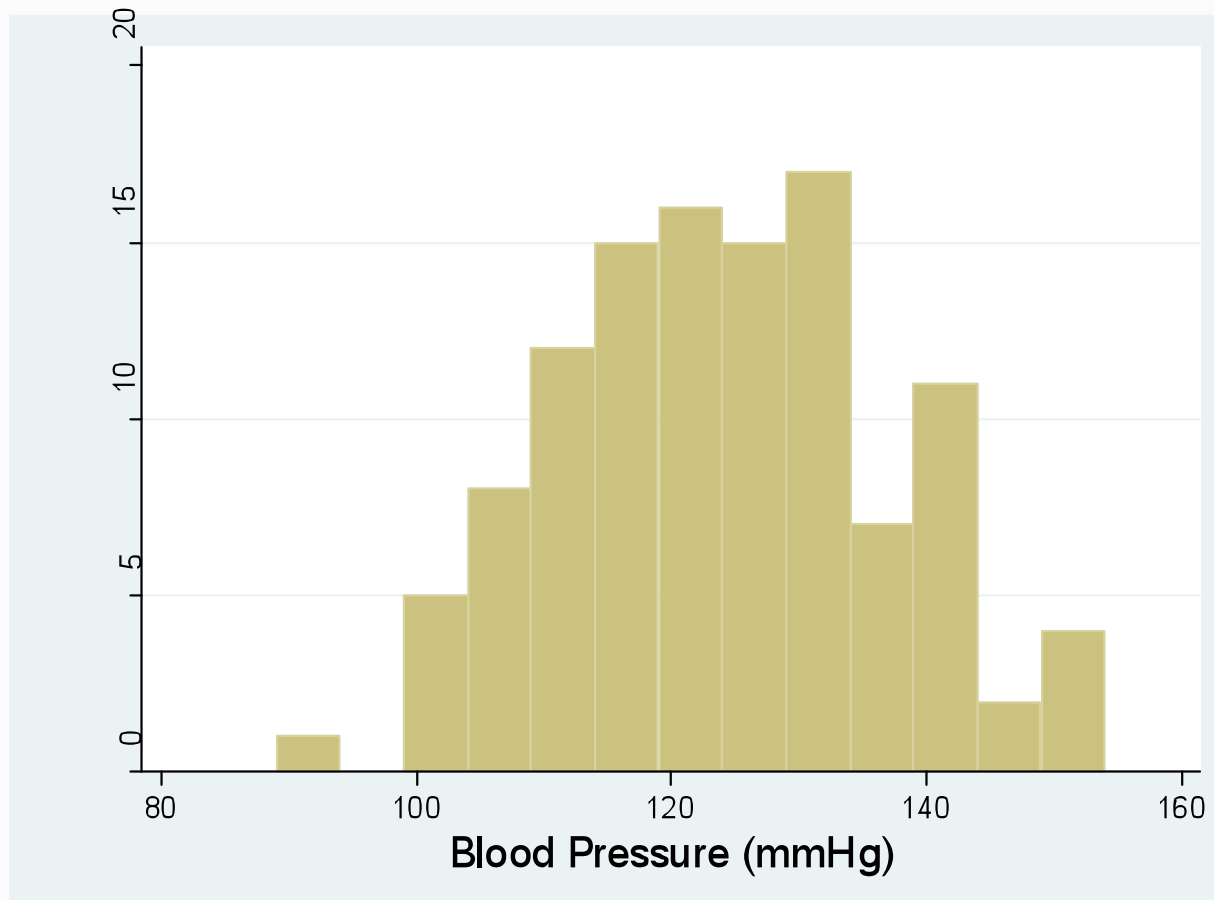




## Pictures of Data: Histograms

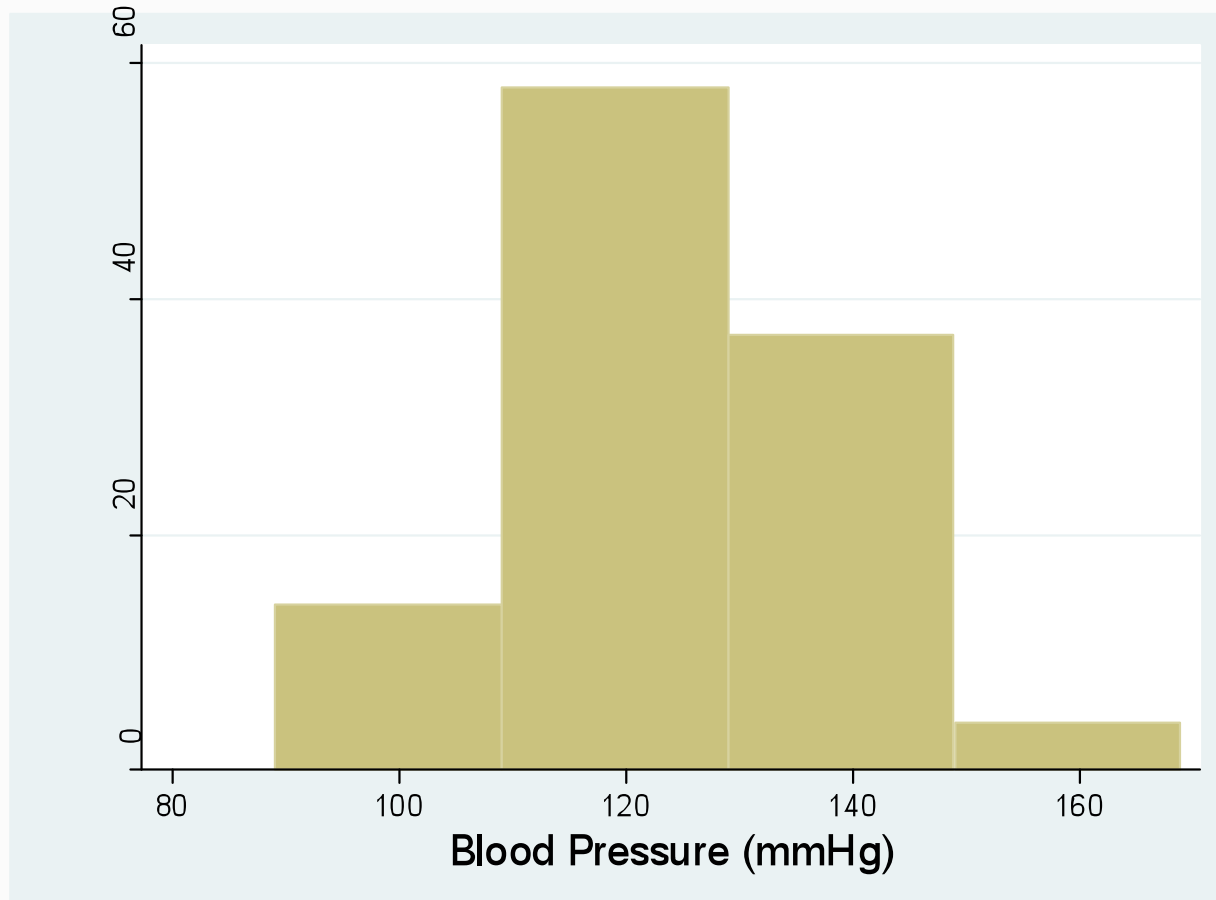
- Suppose we have a sample of blood pressure data on a sample of 113 men
- Sample mean (  $\bar{x}$  ) : 123.6 mmHg
- Sample Median ( $m$ ): 123.0 mmHg
- Sample sd: ( $s$ ): 12.9 mmHg

# Pictures of Data: Histograms



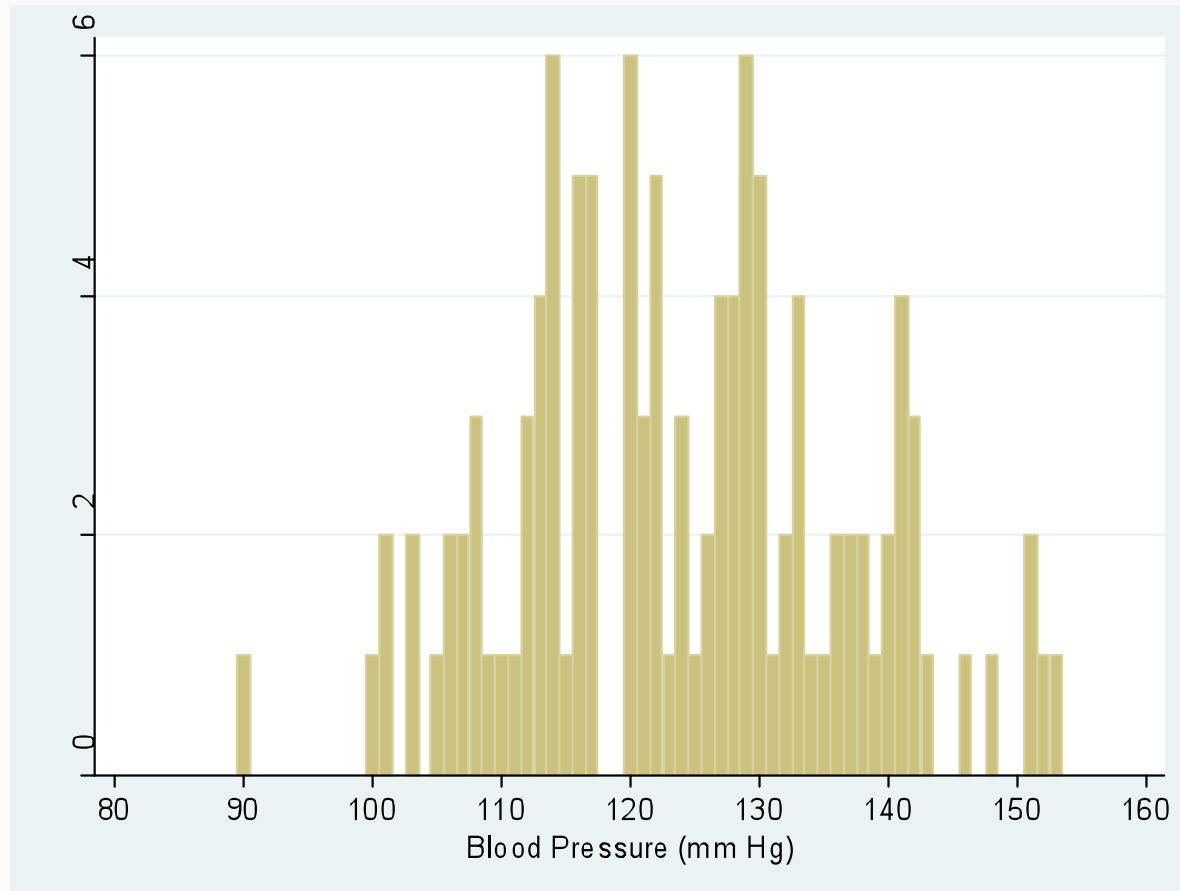
Histogram of the Systolic Blood Pressure for sample of 113 men. Each bar spans a width of five mmHg on the horizontal axis. The height of each bar represents the number of individuals with SBP in that range.

# Pictures of Data: Histograms



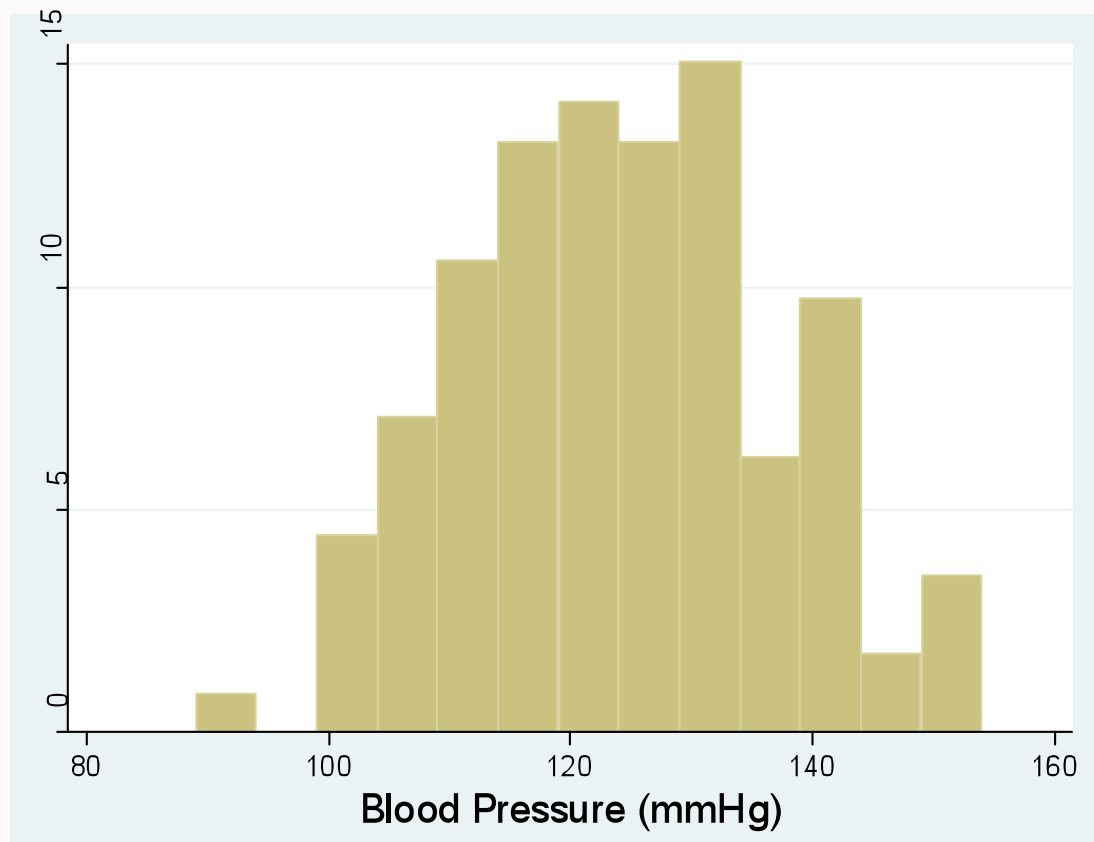
Another histogram of the blood pressure of 113 men. In this graph, each bar has a width of 20 mmHg and there are a total of only four bars making it hard to characterize the distribution of blood pressures in the sample.

# Pictures of Data: Histograms



Yet another histogram of the same BP information on 113 men. Here, the bin width is one mmHg, perhaps giving more detail than is necessary.

## Other Examples



Another way to present the data in a histogram is to label the y-axis with relative frequencies as opposed to counts. The height of each bar represents the percentage of individuals in the sample with BP in that range. The bar heights should add to one.

# Intervals

- How many intervals (bins) should you have in a histogram?
  - There is no perfect answer to this
  - Depends on sample size  $n$
  - Rough rule of thumb: # Intervals  $\approx \sqrt{n}$

<b>n</b>	<b>Number of Intervals</b>
10	About 3
50	About 7
100	About 10