

Descriptive Statistics — Visualization

PSYC 203

Outline

- **Making sense of data**
 - **From qualitative to quantitative**
- Visualizing Data
 - Frequency distributions and histograms
- Language for describing distributions
 - modality, skew, kurtosis
- Why are distributions insufficient?

Summarizing Information/Data

- Imagine we are interested in assessing ‘mathematics knowledge’ of an incoming class of 100 students
- 90 questions on the test and scores are as follows

50	42	43	56	36	48	50	63	58	38
30	49	41	43	61	73	38	39	40	53
51	26	58	52	50	34	80	51	60	63
45	66	58	42	52	67	54	52	47	62
31	50	39	51	36	67	62	45	43	62
49	36	59	39	55	49	50	52	45	50
41	60	49	56	51	32	66	72	63	62
40	39	54	52	42	41	45	62	65	41
41	44	53	60	53	52	88	51	62	67
54	40	51	41	41	42	42	38	51	59

Summarizing Information/Data

- Our goal is to understand (and ultimately communicate) how students performed on this test
- Tough to make sense of data when presented in this 'random' order.
- One way to summarize and make sense is through a frequency table/distribution

Summarizing Information/Data: Ungrouped Frequency Distributions

- Lists all possible values and the number of times each occurs
- Steps
 - Identify largest and smallest observed values
 - List all values between the largest and smallest
 - Tally the number of times each score was observed
 - Report the full table
- Only **one** possible solution in a given set of data

Summarizing Information/Data: Ungrouped Frequency Distributions

- **R code**
- What do you see as positive and negative aspects of this approach?

exam	Freq
26	1
30	1
31	1
32	1
34	1
36	3
38	3
39	4
40	3
41	7
42	5
43	3
44	1
45	4
47	1
48	1
49	4
50	6
51	7
52	6
53	3
54	3
55	1
56	2
58	3
59	2
60	3
61	1
62	6
63	3
65	1
66	2
67	3
72	1
73	1
80	1
88	1

Summarizing Information/Data: Grouped Frequency Distributions

- Rather than reporting individual scores, we cluster the original scores into class intervals.
- How do we construct the intervals?
 - Intervals should be mutually exclusive and exhaustive
 - Number of intervals
 - Rule of thumb is 10 to 20 (might be less with small sample size)
- Width of intervals
 - Should be the same
 - Logical (2, 5, 10, etc.)

Summarizing Information/Data: Grouped Frequency Distributions

- **R code**
- What do you see as positive and negative aspects of this approach?

	scores	Freq
1	(25.5,35.5]	5
2	(35.5,45.5]	33
3	(45.5,55.5]	32
4	(55.5,65.5]	21
5	(65.5,75.5]	7
6	(75.5,85.5]	1
7	(85.5,95.5]	1

	freq	cumulativefreq
[25.5,35.5)	5	5
[35.5,45.5)	33	38
[45.5,55.5)	32	70
[55.5,65.5)	21	91
[65.5,75.5)	7	98
[75.5,85.5)	1	99
[85.5,95.5)	1	100

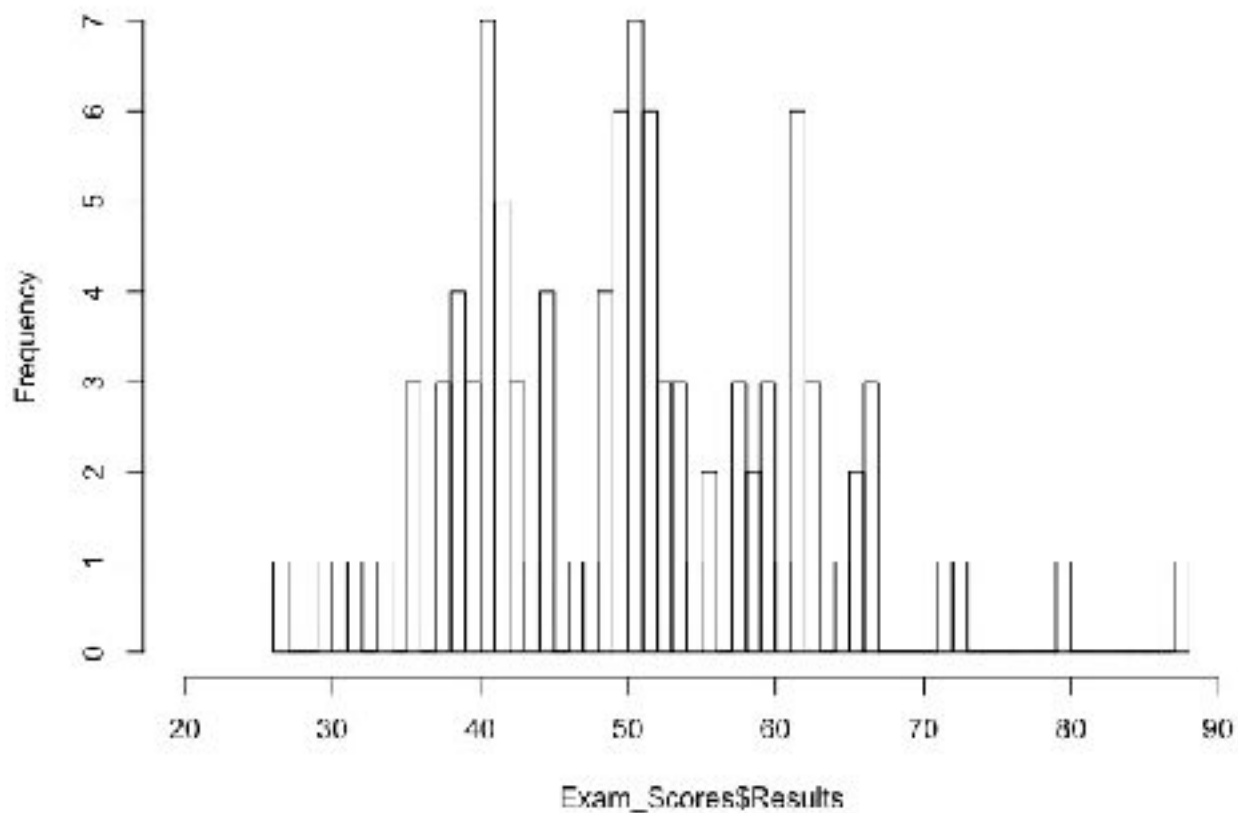
Summarizing Information/Data: Histograms

- Graphic representation of the information presented in a frequency table.
- Can be either ungrouped or grouped
 - Different statistical programs use different defaults so pay attention

Summarizing Information/Data: Histograms

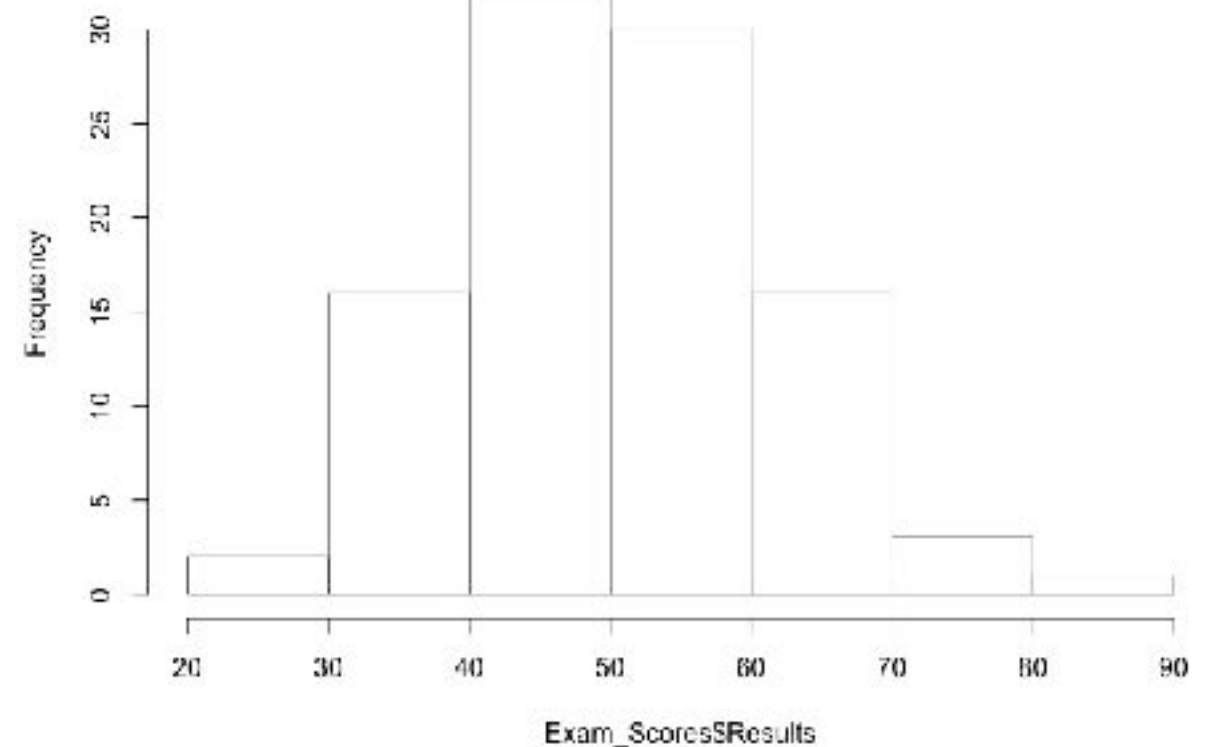
R code

Histogram of Exam_Scores\$Results



Ungrouped

Histogram of Exam_Scores\$Results



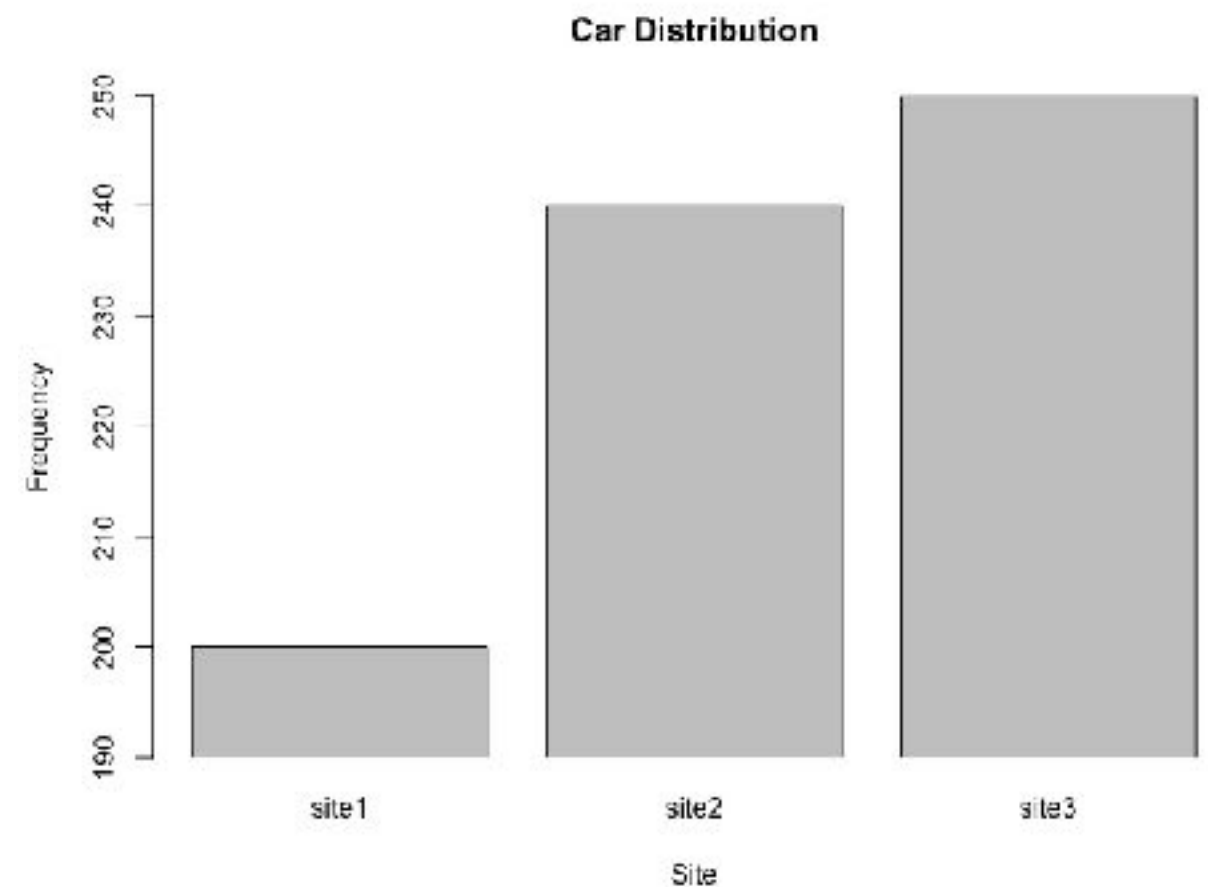
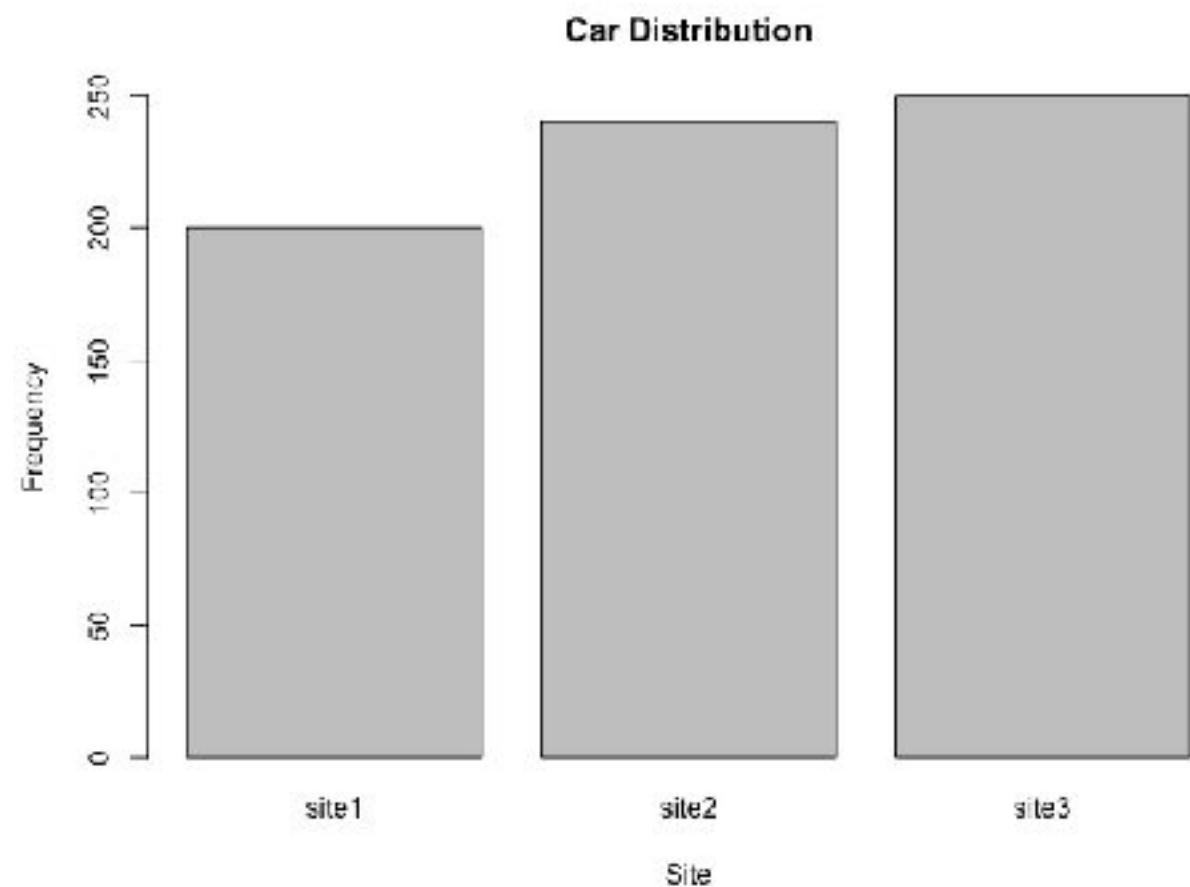
Grouped

Summarizing Information/Data: Histograms Tips

- Values on the horizontal axis are often not in their 'true' position relative to zero.
- Can influence how data are interpreted

Same Data, Different Impression

- **R code**

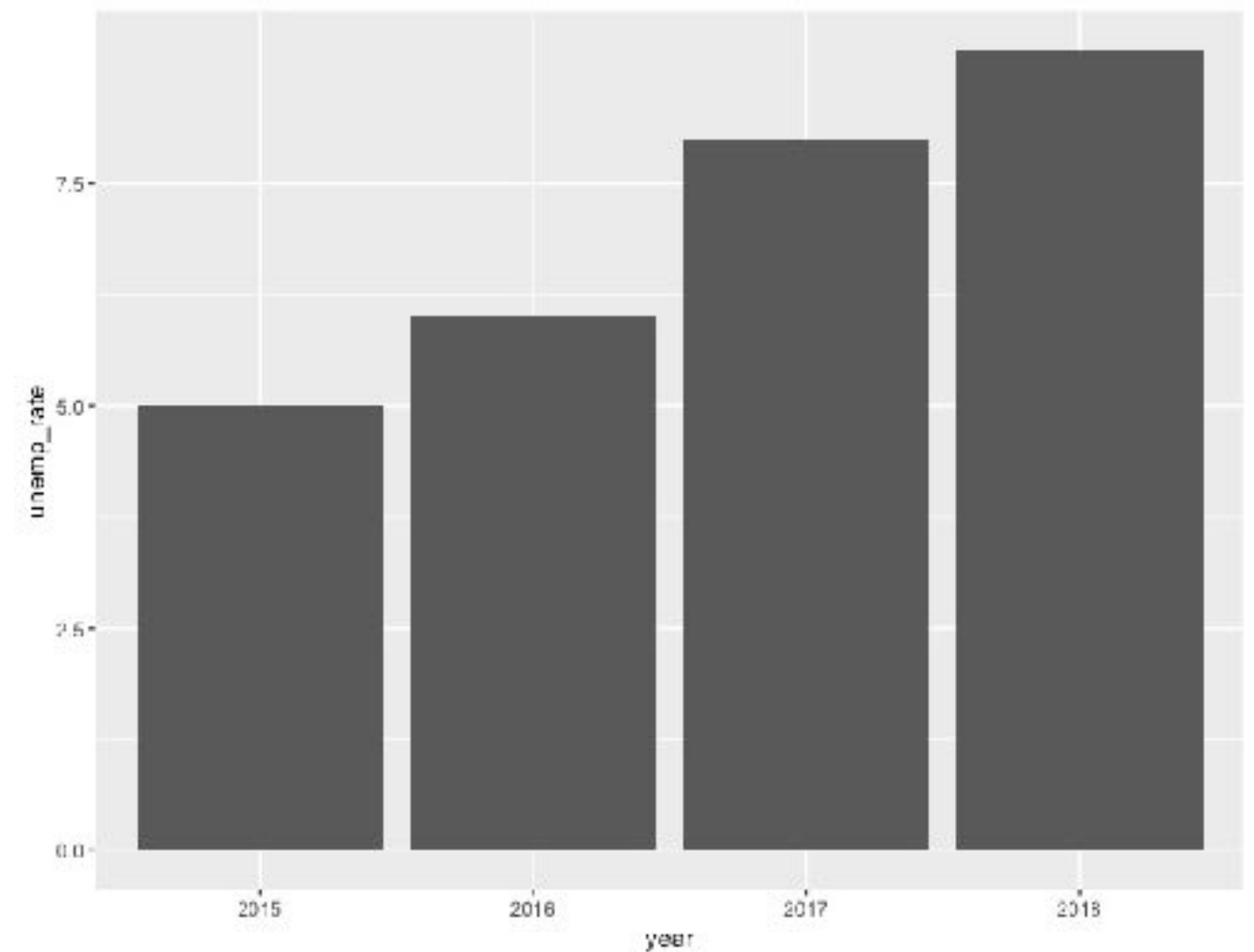
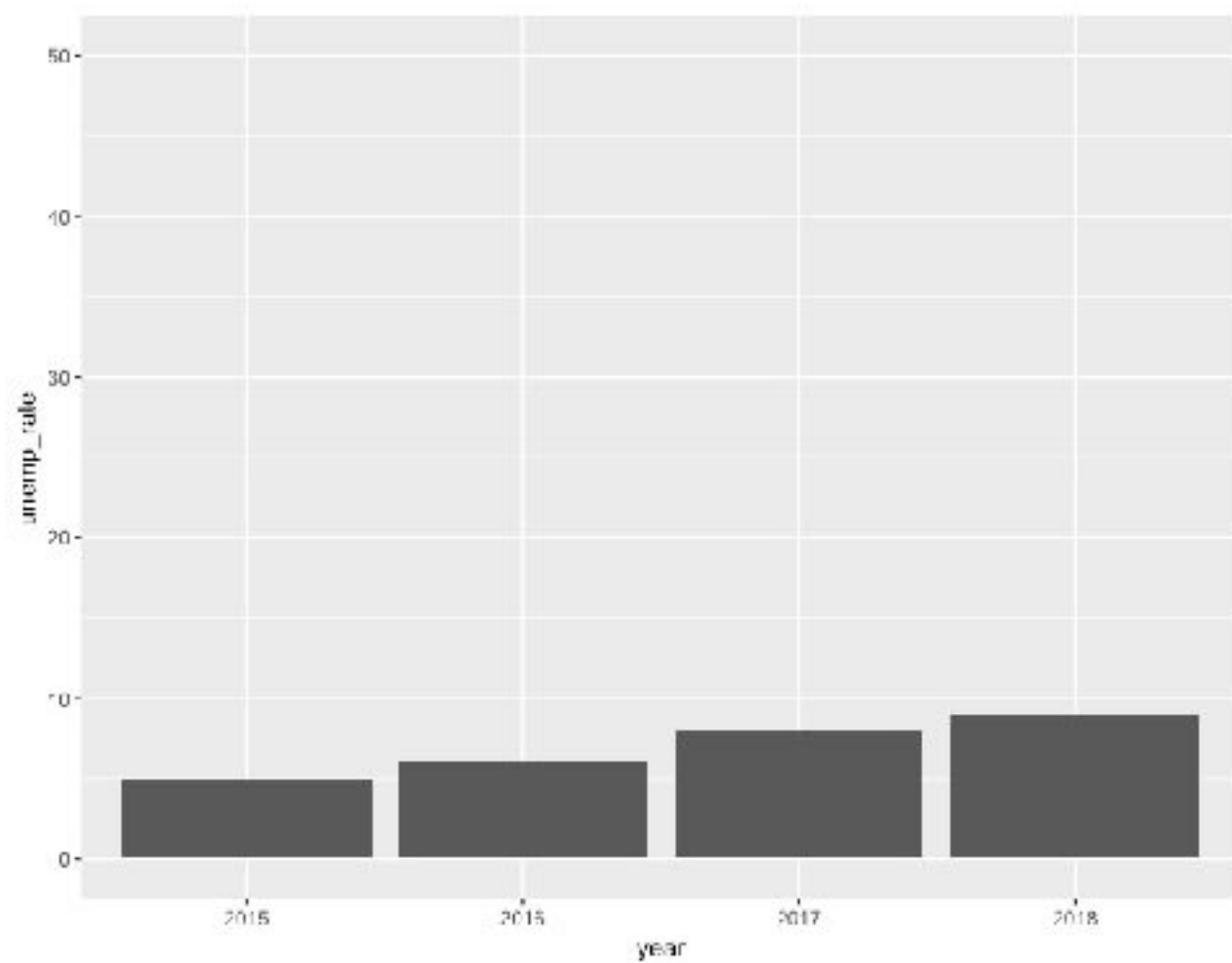


Summarizing Information/Data: Histograms Tips

- Values on the horizontal axis are often not in their 'true' position relative to zero.
 - Can influence how data are interpreted
- Height of histogram relative to width is important

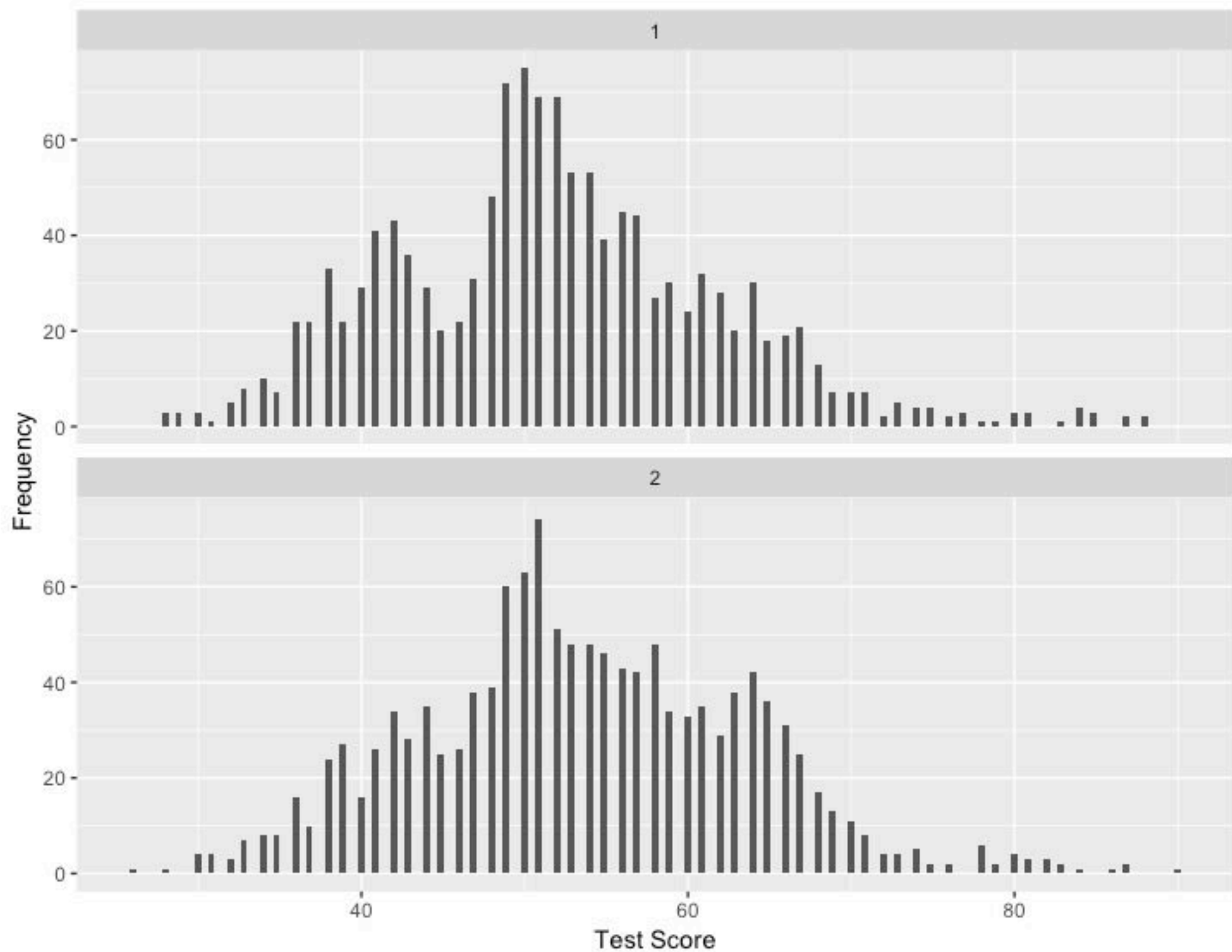
Who's Telling the Story?

- **R code**



Comparing Histograms

- **R code**



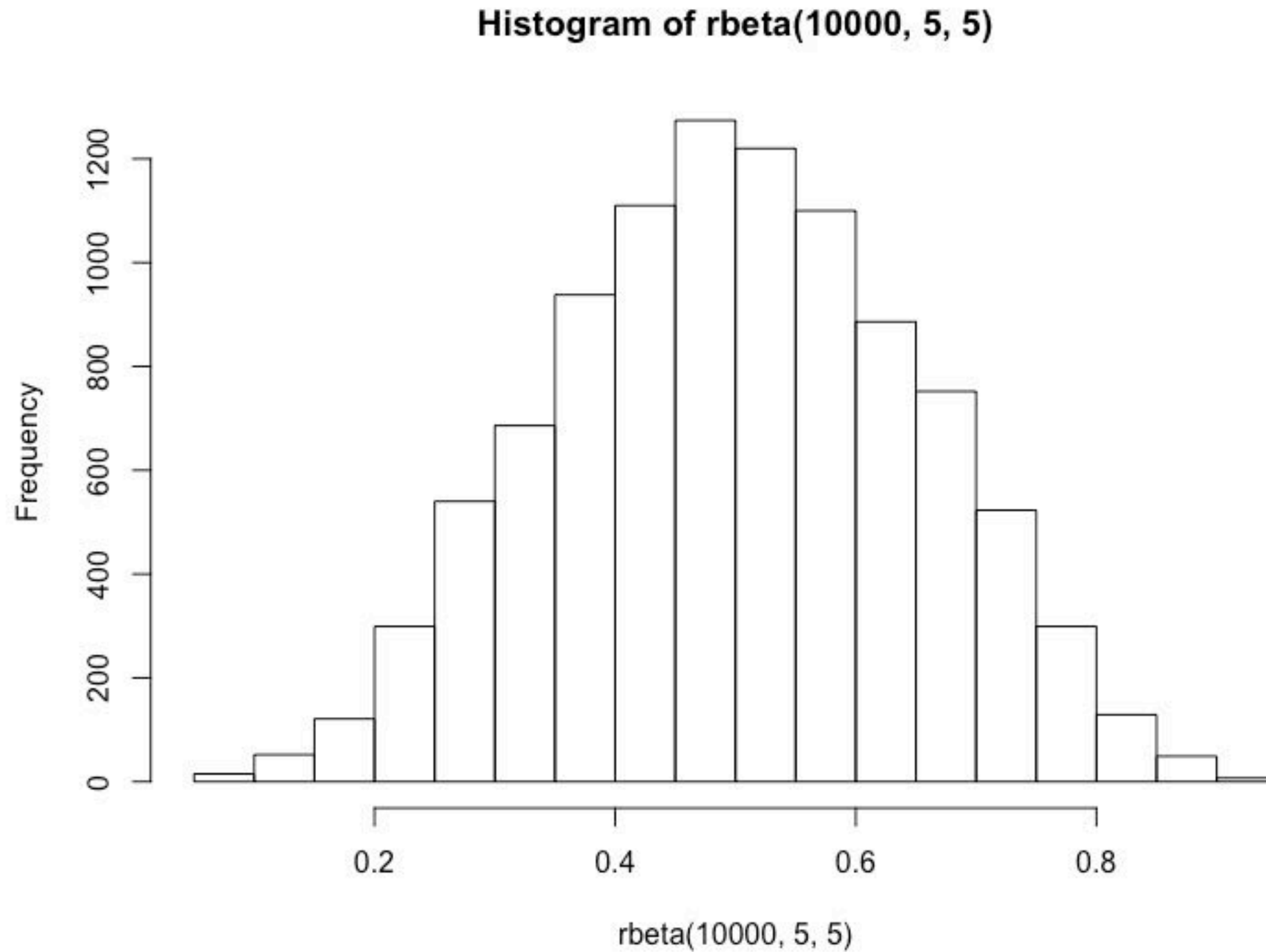
Outline

- Making sense of data
 - From qualitative to quantitative
- Visualizing Data
 - Frequency distributions and histograms
- **Language for describing distributions**
 - **modality, skew, kurtosis**
- Why are distributions insufficient?

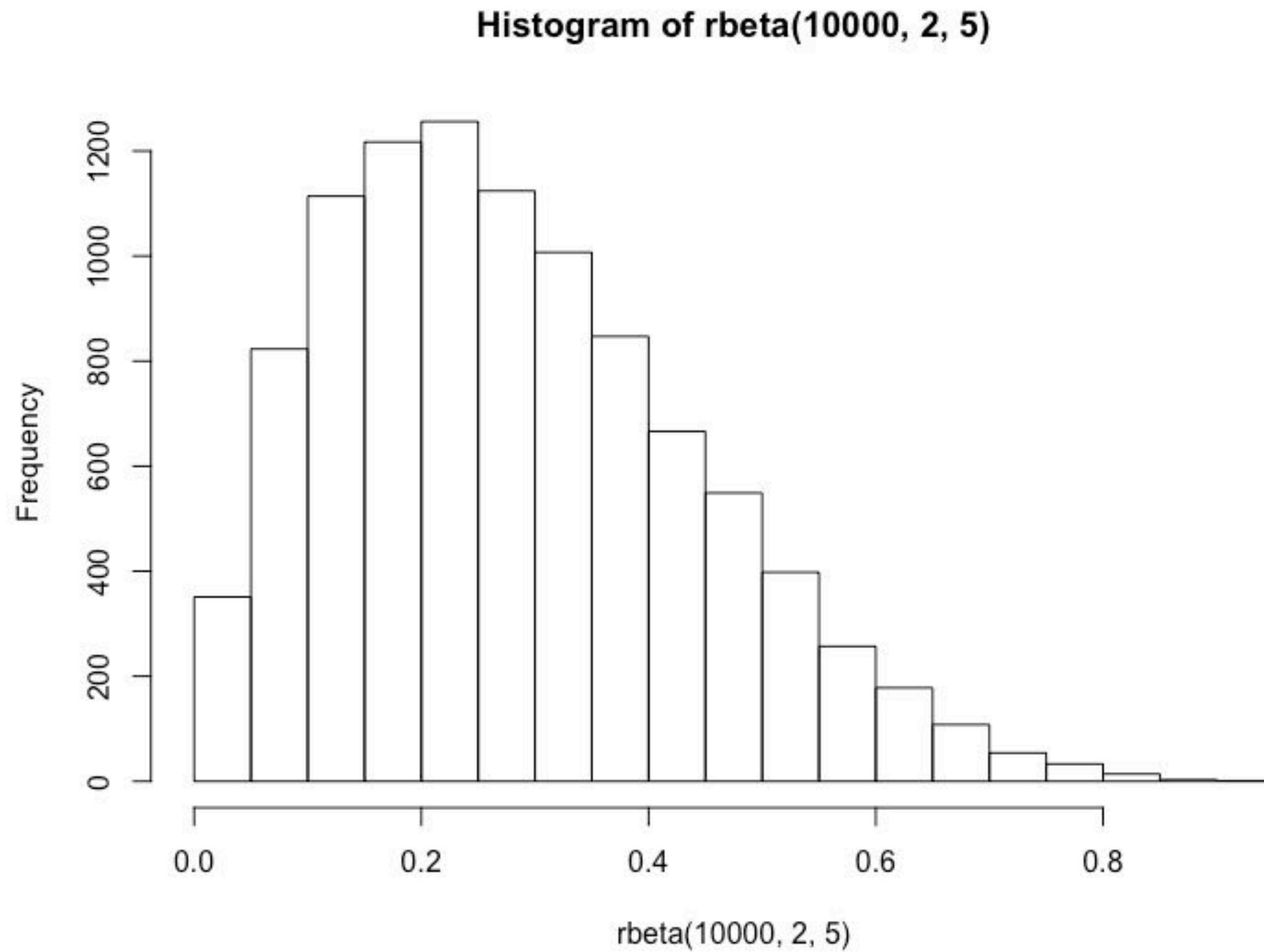
Interpreting Distributions

- How many peaks are present?
 - modality (unimodal, bimodal)
- Is the distribution symmetric?
 - skewness (positive, negative)
- How are scores concentrated?
 - kurtosis (mesokurtosis, platykurtosis, leptokurtosis)

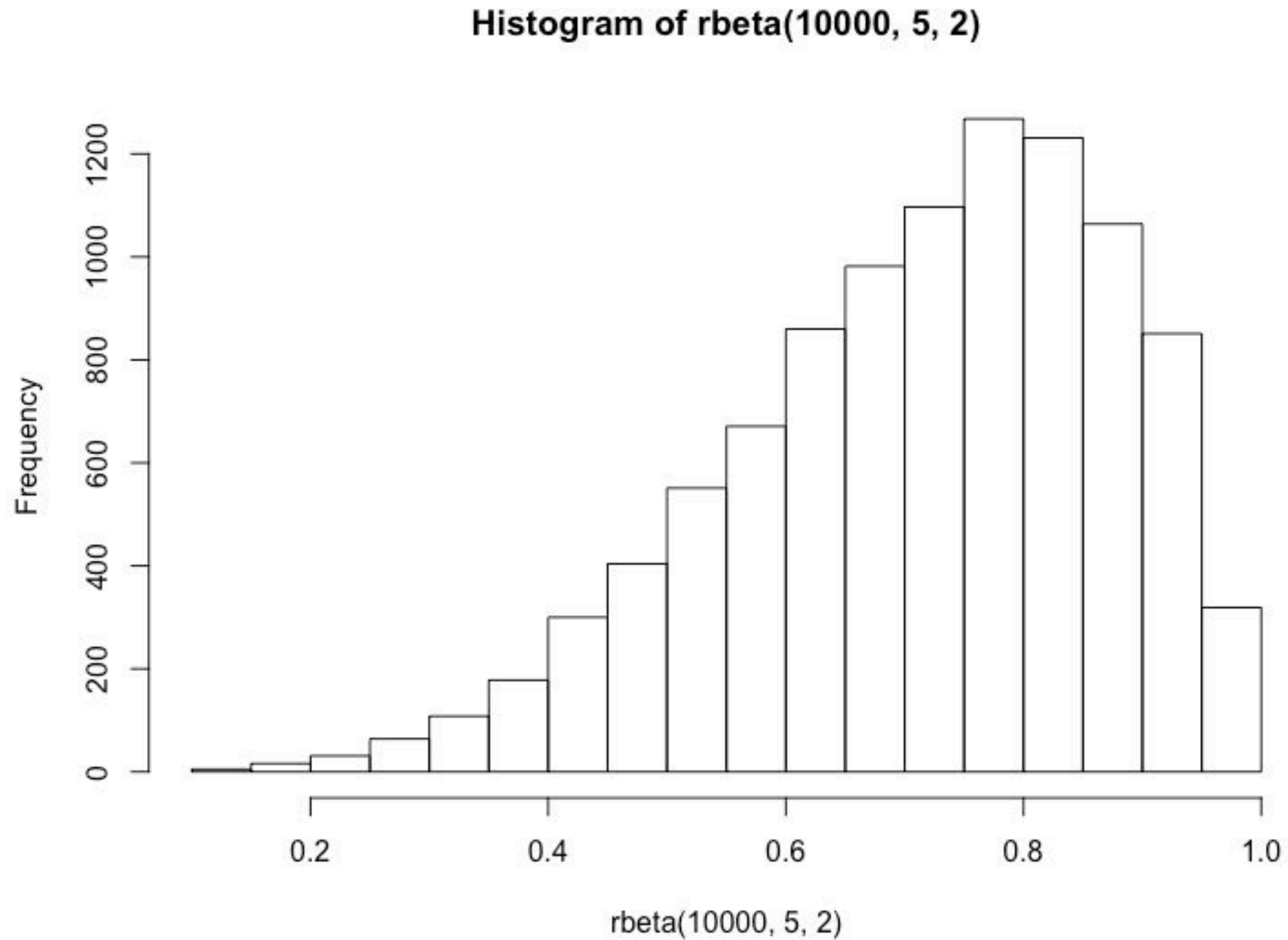
Distribution #1



Distribution #2



Distribution #3



Outline

- Making sense of data
 - From qualitative to quantitative
- Visualizing Data
 - Frequency distributions and histograms
- Language for describing distributions
 - modality, skew, kurtosis
- **Why are distributions insufficient?**

Why Are Distributions Insufficient?

- We typically want to talk about central tendency and dispersion
- Subjective v. Objective summary