Exploring “What’s Typical?” Using Graphical Displays of Data

education.ti.com/tiwebinars

Agenda

- Welcome & Introductions
- What is Typical and What is Not
  - Using Data to Make Decisions
  - Making Decisions in Context
  - Comparing Distributions
  - Communicating Results
  - Misconceptions with graphical representations
  - Cautions about real data

Expected Outcomes:

- **Explore** graphical representations of data sets, thinking about shape, center and variability

- **Learn** how TI technology can give students meaningful opportunities to develop an understanding of what is typical and what is not for a given set of data

Analyzing Data

The story in the data- what is typical and what is not:

Shape, Center and Variability

Adapted from Kader & Memer, 2008

Table 3: Effect of exercise on heart rate

	Rest	Anticipation	Difference	Exercise	Difference
Henry	72	76	4	87	15
Megan	116	120	4	175	59
Laura	79	84	5	96	17
David	97	99	2	100	3
John	90	93	3	176	86
Michael	67	75	8	132	65
Sarah	115	116	1	176	61
Claire	82	83	1	141	59
Rosanna	95	98	3	113	18
Rachel	82	87	5	136	54
Hattie	77	82	5	96	19
Rosie	105	110	5	153	48
Alex	79	82	3	90	11
Katheryn	99	102	3	152	53
Rebecca	82	89	7	156	74
Siobhan	87	94	7	170	83
Mark	82	94	12	128	46
Thomas	98	76	-22	172	74
Lougie	80	92	12	132	52
Richard	95	115	20	141	46
Edward	79	128	49	144	65
Jamie	80	76	-4	100	20
Jacqueline	85	88	3	172	87
Hope	91	130	39	177	86
Holly	83	89	6	168	85
Emma	77	81	4	165	88
James	87	87	0	111	24
Seamus	84	92	8	122	38
Ciaran	93	82	-11	185	92
Ryan	87	92	5	170	83
Christine	91	93	2	124	33
Anne	90	94	4	176	86
Mark	100	101	1	160	60
Alistair	84	85	1	166	82
Anthony	64	108	44	167	103
Patrick	72	88	16	151	79
Tom	84	91	7	147	63
Douglas	86	97	11	173	87

The data: British children ages 9 to 11

Robson, L. (2016). The Human Body – a Lean Mean Exercise Machine, The Heart Teachers notes. Department of Biomedical Science, University of Sheffield www.coursehero.com/file/16351710/Heart-teachers-notes/

The number of words a class of 32 students could read at the end of the second grade is shown in the dot plot below. Are any of the students extremely poor readers or extremely good readers for their grade level? Explain how you know.

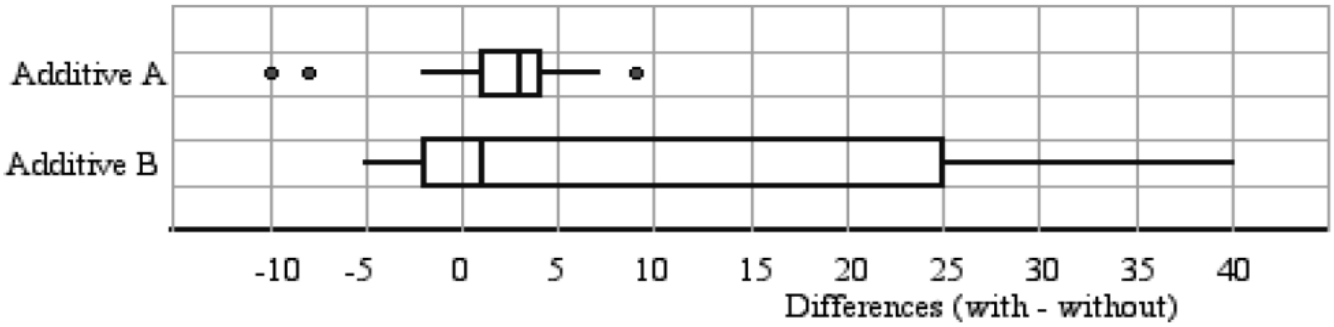
Working with Box Plots

Using problems from the Advanced Placement Exam

2005 #1 Gasoline Additives

2015 #1 Salary Increases

What students were asked……



(b) Two ways that the effectiveness of a gasoline additive can be evaluated are by looking at either

- the proportion of cars that have increased gas mileage when the additive is used in those cars

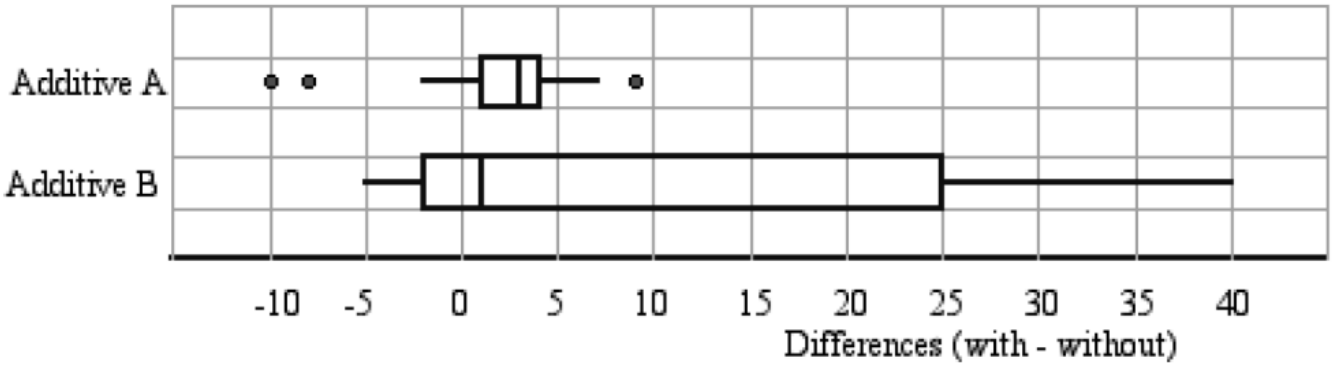
  or

- the mean increase in gas mileage when the additive is used in those cars.

i. Which additive, A or B, would you recommend if the goal is to increase gas mileage in the highest proportion of cars? Explain your choice.

ii. Which additive, A or B, would you recommend if the goal is to have the highest mean increase in gas mileage? Explain your choice.

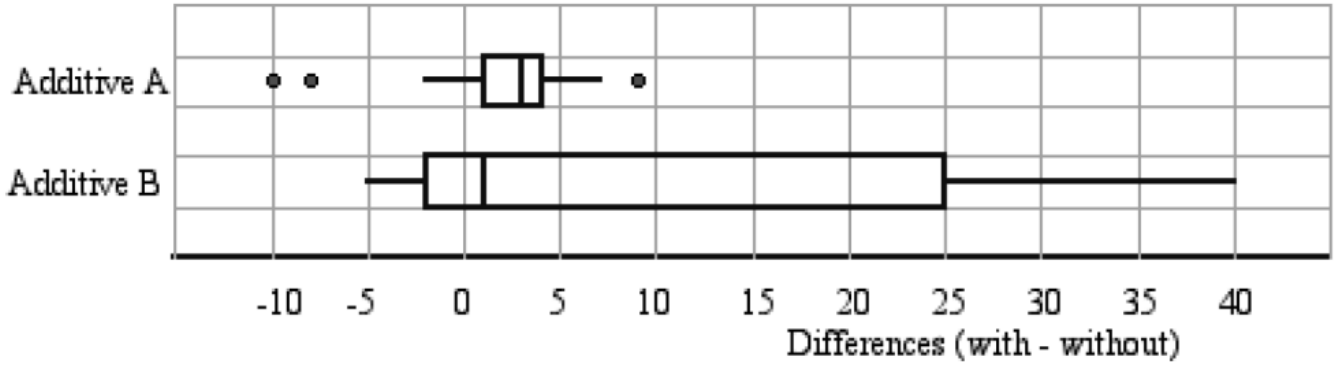
The model solution:



Part (b i):

Additive A is better at increasing the mileage in the greatest number of cars. The mileage increased for at least seventy-five percent of the cars with additive A, whereas the mileage decreased for more than twenty-five percent of the cars with additive B.

The model solution:



Part (b ii):

Additive B appears to produce a higher mean mileage gain than additive A. The boxplot for additive B clearly shows that the distribution of differences is skewed to the right, which will pull the average towards the larger values. The mean difference for additive B will be substantially greater than the median of 1. On the other hand, the distribution of differences for additive A has much less variability, as seen by comparing the lengths of the two boxes, and appears to be skewed to the left. The mean difference for additive A will be less than the median of 3.

Gasoline Additives

- From the chief reader notes posted on AP Central: "What was the intent of this question? The intent of this question was to assess students' ability to use summary information for two distributions (one for each additive) to construct parallel boxplots. Students were asked to check for outliers and to identify any outliers on their graph. ***Finally, students were asked to interpret the information provided in the boxplots by making recommendations based on two different goals. "***
- Many common misconceptions from students occur when students make a choice, and are not able to adequately articulate why they are making that choice and at the same time why they did not choose the other. (In Part (b ii) some students described the distribution of B but failed to make the comparison to the distribution of A.)
- Recommendations from the chief reader: "Students need to look carefully at the information they are given before making appropriate graphical displays. Identifying outliers and describing the impact of the shape of the distribution on basic statistics are important aspects of exploratory data analysis in which students could use more practice. Exposing students to practical problems where different goals or priorities lead to different answers or results would be extremely beneficial."

Salary Increases

(b) Suppose both corporations offered you a job for $36,000 a year as an entry-level accountant.

(i) Based on the boxplots, give one reason why you might choose to accept the job at corporation A.

(ii) Based on the boxplots, give one reason why you might choose to accept the job at corporation B.

The model solution:

(i) Five years after starting, at least 3 out of 30 (10%) of the salaries at Corporation A are greater than the maximum salary at Corporation B. If I accept the offer from Corporation A, I might be able to make a higher salary at Corporation A than at Corporation B.

The model solution:

(ii) Five years after starting, the minimum salary at Corporation B is greater than at Corporation A. In fact, at Corporation A it looks like some people are still making the starting salary of $36,000 and never received a raise in the five years since they were hired. So if I work at Corporation A, I might never receive a raise in salary.

Salaries

What was the intent of this question?

The primary goals of this question were to assess a student's ability to (1) compare features of two distributions of data displayed in boxplots and (2) identify statistical measures that are important in making decisions based on data sets.

Part (a)

- Many students did not communicate clearly.
- Some students tried to determine the shape or type of distribution from the boxplots.
- Some students described the two boxplots but did not use comparative language.
- Some students omitted the context; students need to clearly identify the variable of interest in their responses.
- Some students did not understand which descriptive statistics can be determined (and compared) from a boxplot and which cannot. For instance, medians can be compared but means cannot. (Only five-number summaries and outliers can be determined from boxplots.)

Part (b)

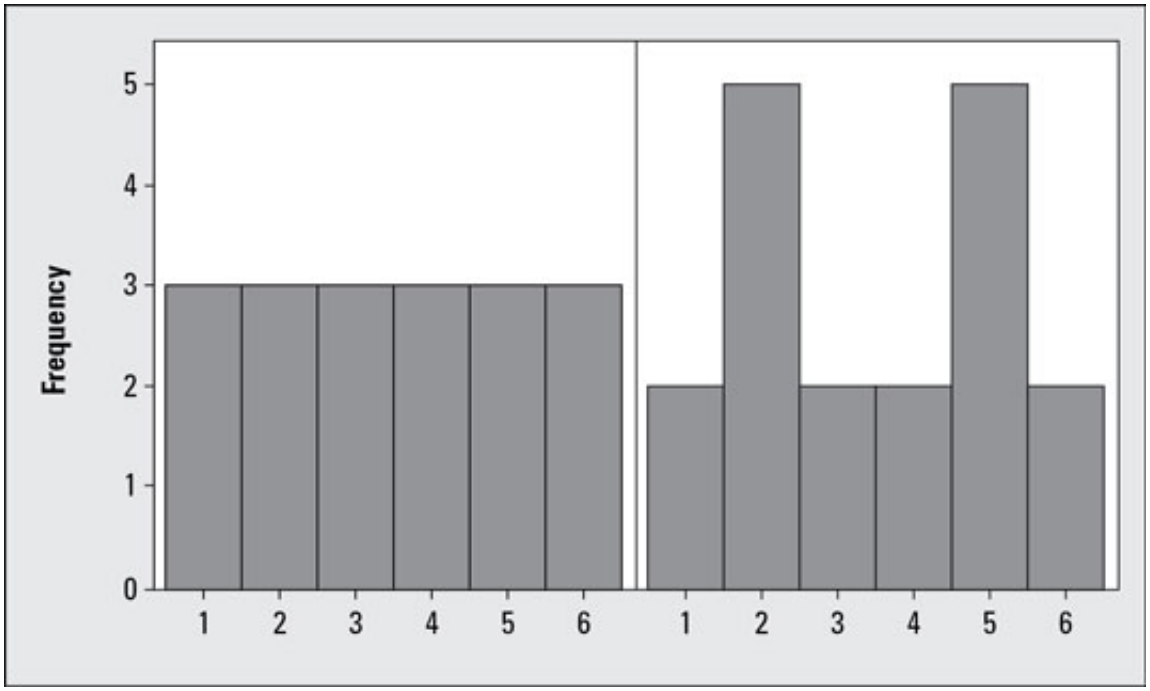
- Some students used statistical language when trying to convey a non-statistical meaning.
- Some students did not understand the five-number summary, for instance thinking that a wider interval means more people were in it.
- Some students referred to the end of the whisker as the maximum when there were outliers.

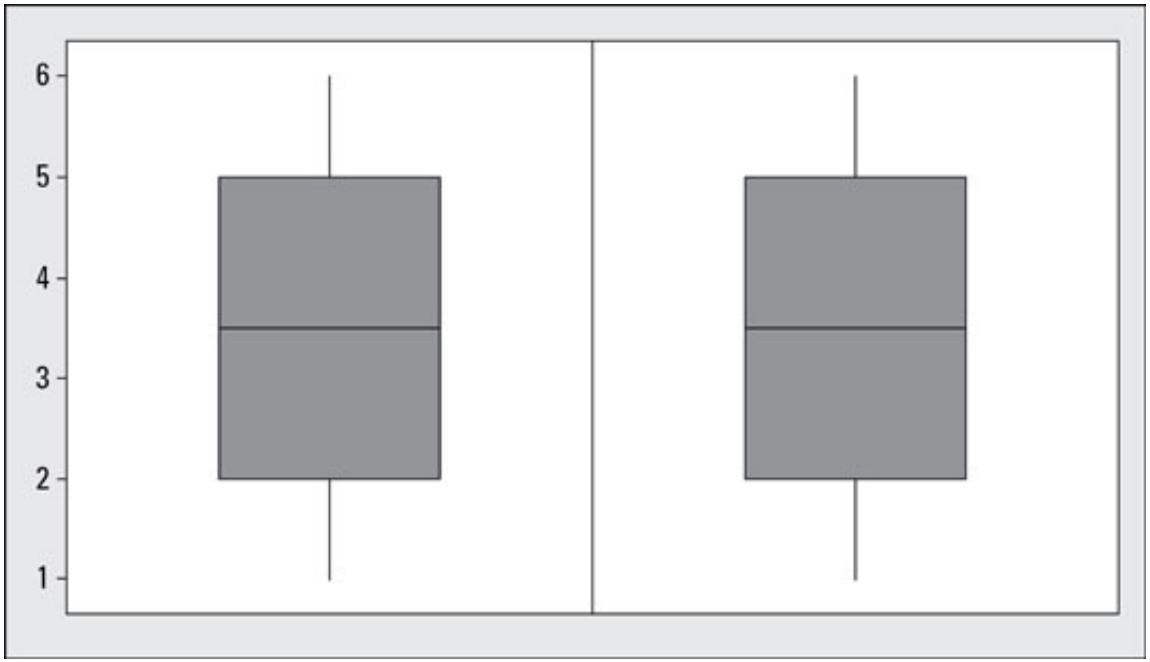
one

Salaries

Teachers should discuss what can be learned from boxplots and what cannot. Means and standard deviations cannot be determined from boxplots, and in many cases it is not possible to determine which of two boxplots represents data with a higher mean or standard deviation. Similarly, complete shape information cannot be determined from boxplots. The only definitive measures that can be learned from boxplots are the elements of a five-number summary and outliers that are extreme enough to be illustrated in the boxplot.

Teachers should provide a lot of practice with interpreting data from a variety of contexts, and with writing clear explanations that involve the context. In particular, the context should be incorporated along with statistical information when making conclusions.





Misconceptions

- Misuse “normal” in describing shape
- Do not view spread from the mean or median
- Confuse variability with frequency
- Perform the calculations but cannot connect the mean and spread to a distribution (Chance et al, 2004; Goth & Bergner, 2006).
- Confuse data values and frequencies
- Lose reference to the context (Cooper & Shore, 2008)
- See distribution as individual values (Hancock, Kaput, & Goldsmith, 1992; Konold et al, 2004; Ben-Zvi & Arcavi, 2001).
- Can not read between the data to summarize and compare nor beyond the data to make a prediction or inference. (Curcio, 1987)

Misconceptions

- Confuse spread in an interval in boxplot with density
  (Konold et al, 2004)
- Can determine shape of distribution from boxplot
- Confuse outliers with maximum or minimum
- Describe a value as “in the lower quartile”

Gender	Age	Handedness	Height (cm)	Foot length (cm)	Arm span (cm)	Favorite subject	Languages spoken	Texts sent yesterday
Male	17	Left	175	24.75	170	Science	2	59
Male	15	Right	173	27	172	Math/stat	1	1
Female	17	Right	175	26.5	178	Science	1	20
Female	16	Right	161	27	162	Math/stat	1	10
Male	18	Right	178	27	169	Phy ED	2	70
Female	18	Right	158	21	164	Art	2	50
Female	17	Right	163	23.5	168	English	1	61
Male	16	Right	169	26	174	Comp Sci	1	70
Female	17	Right	162.6		157.5	Music	2	45
Female	18	Right	166.5	26	170	History	1	100
Male	17	Right	193.4	27.8	189.4	Comp Sci	1	26
Male	16	Left	177	67	128	Comp Sci	1	18
Female	17	Right	154	23	156	Other	1	200
Female	17	Right	148	22.5	149	Comp Sci	2	120
Male	16	Right	183	25	192	Music	2	12
Female	16	Right	165	22	156	Science	1	200
Female	17	Left	154.4	23	124	Science	2	45
Female	16	Right	179	26	174	History	1	16
Female	16	Right	174	25	171	History	2	36
Female	16	Right	158	20.5	165	English	1	15
Female	16	Right	171	25	168	Music	2	32
Female	17	Left	169	24	175	Science	1	25
Female	16	Right	166	25	168	English	1	0
Female	18	Right	170	23	156	Art	1	5

Table 1: Random Sample of Grade 11 students from the United States from Census at School American Statistical Association

Real data: What is typical and what is not?

CENSUS at SCHOOL

- Home
- Student Section
- Teacher Section
- Random Sampler
- International

Welcome to Census at School - United States

Census at School is an international classroom project that engages students in grades 4-12 in statistical problemsolving. Students complete a brief online survey, analyze their class census results, and compare their class with random samples of students in the United States and other countries. More

What's New?

Census at School New Zealand now hosts the random sampler for the international Census at School data, the New Zealand data, and also for the cleaned USA data. Their online random sampler allows students and teachers to take random samples up to size 1000 from the international, New Zealand, or U.S. database and either download the data or start up the free, online iNZight Lite software with the data already loaded and ready for analysis. The international database includes data from Australia, Canada, New Zealand, the United Kingdom, and the United States. Census at School - United States also provides a Random Sampler tool to generate a sample of USA Census at School data. This provides access to the messy U.S. data that has not been cleaned. For more information, see the article in *Statistics Teacher*.

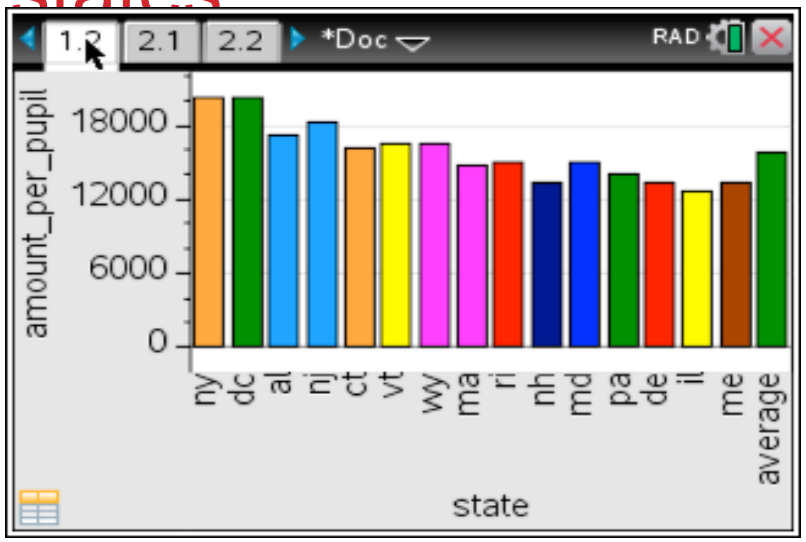
The American Statistical Association and Population Association of America are seeking champions to expand U.S. Census at School nationally. Be in on the ground floor and get involved today.

About Census at School

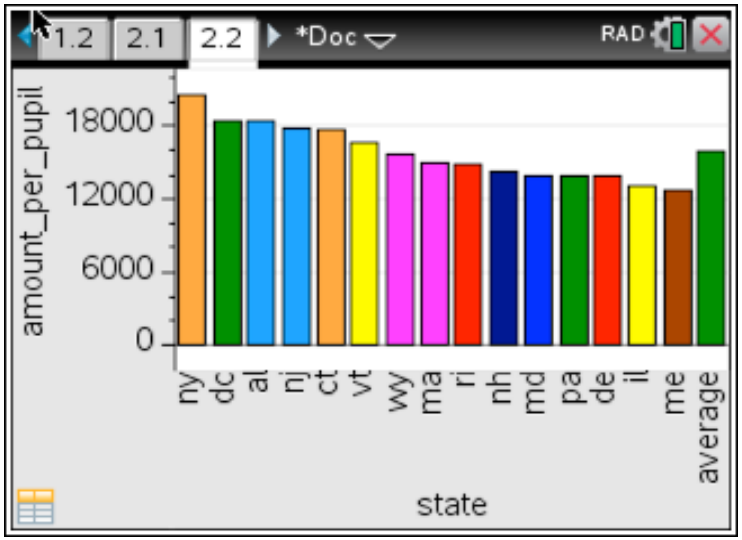
Privacy Statement

Resources

Change in dollars spent per pupil in selected states

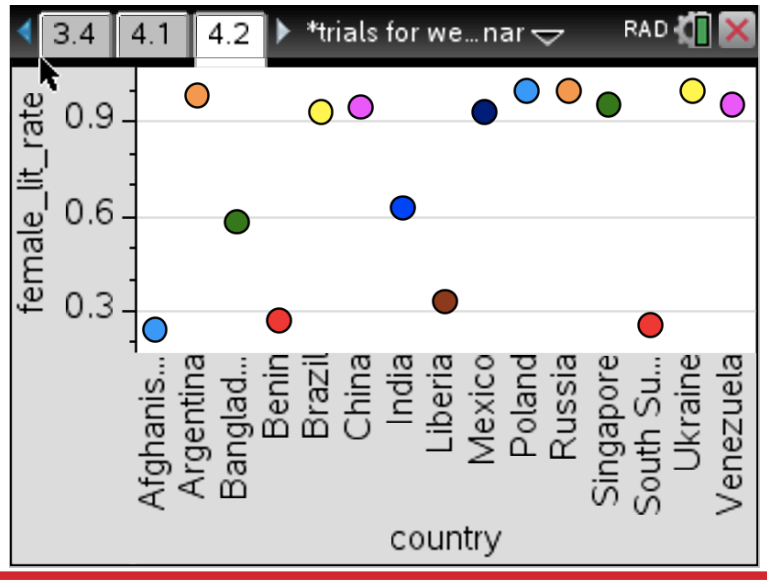


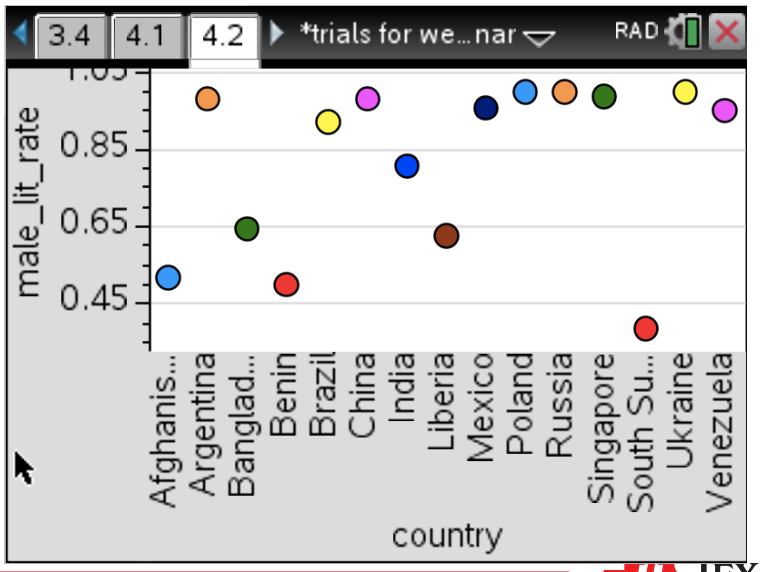
2010



2014

Comparing literacy rate in selected countries





Suggestion:

- “make that first week or so not about the topics being reviewed, but about the critical component of communication. Even if kids can already make a histogram or find a median, they've never had to write explanations, interpretations, and conclusions that were clear, complete, concise, and in context. You can set an important tone for the rest of the year by emphasizing communication. I'll conjecture that at the end of the year students lose more credit on the AP Exam for poor writing (and the hazy grasp of concepts that may underlie that) than for errors in statistical methodology.”
  David Bock, AP list serve

References

- Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics, 45*, 35-65.
- Chance, B., delMas, R. C., and Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295-323). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Cooper, L., & Shore, F. (2008). Students' misconceptions in interpreting center and variability of data represented via histograms and stem-and-leaf plots. *Journal of Statistics Education* Volume 16, Number 2 (2008)
- Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education, 32*(2), 124-158.
- Groth, R., & Bergner, J. (2006). Preservice elementary teachers' conceptual and procedural knowledge of mean, median, and mode. *Mathematical Thinking and Learning,* 8(1), 37-63.
- Hancock, C., Kaput, J. J., & Goldsmith, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist*, *27*(3), 337-364.
- Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2004). *Data seen through different lenses.* Unpublished manuscript, Amherst, MA: University of