Histogram exercise
Who got it?
What are the percentages?
What’s “wrong” with this histogram?
What are the ingredients of a good graph?

Homework:
Use any appropriate means, but be clear and complete.
   Word processed best (copy and paste graphics)
   Hand written OK IF LEGIBLE AND CLEAR
   EXCEL OK for graphics but NOT for stats
   Reference graphics in your interpretation, (Figure 1) and have graphics labeled appropriately.
Many problems (AND ALL EXAMS) require minitab. Practice.
   All College network computers: Start > Kenyon Programs > Mathematics
   www.minitab.com
   rent or buy at www.e-academy.com/minitab (30 bucks/ 5 months)
Grading rubric online in course materials
   Understanding, Analysis Plan, Solution, Presentation (0-3 points each)
   Not all categories apply to each problem.
   Problems that require more are worth more.
   Homework graded by student grader. Reports and exams graded by me.

Today: Descriptive statistics (are we talking about Q or C variables?)

Histogram from Wednesday (any)
   How can we describe this data?
   How can we quantify this description?

Overall shape, center, and spread
Overall shape:
   1. how many modes (humps) in the distribution?
   2. symmetric or skewed distribution?
   3. any outliers apparent?

“But what is a skewed distribution? What are outliers?”
   Skew: asymmetry of distribution – one “tail” longer than the other. Distribution is said to be skewed in the direction of its conspicuously longer “tail.” (NOT in the direction of the “body”.) Can be quantified – but not in this course.
   Outliers: “extreme” observations. Of concern for two primary reasons:
   1. May be unrealistic, spurious
   2. May have undo influence on statistical summaries, inferences
      For example: Mean, Standard Deviation (see below)
Center and spread – 2 ways to go after these descriptions
Ranks – median and IQR
Sums of values – mean and standard deviation

Measuring central tendency (How can we do it?)

Mean – Take the average. Review – summation and division.
   Our classmate whose hometown has “A Trillion” people…
   37,037,215,037 (more than six times the global population) vs. 184,846.15
Trimmed Mean – removes extreme values before summing

Median – Find the middle value. Hometowns: 20,000 people

When will the mean and the median be approximately equal? When will they be different?

Measuring spread in the data (How can we do it?)

Range – Maximum – Minimum observations

Interquartile range (IQR) – 3rd – 1st Quartiles
   1st quartile: middle value of the values below the median
   3rd quartile: middle value of the values above the median
   Quartile: 4, Percentile: 100, Decile: 10 (all even divisions of a data distribution).
   You get the idea…

1.5*IQR rule for classifying outliers (more of a guideline, actually)

Sums of squares method: Variance and Standard Deviation
   Deviation: any observation minus the mean observation
   Deviations sum to zero – How can this be avoided?
   Sum of squares (SS): sum all of the squared deviations.
   Since the deviations sum to zero only N-1 are independent.
   Variance (s² or σ²): Representative Squared Deviation SS/N-1.
   Standard Deviation (s or σ): Same units as original variable sqrt(s²).

What are the units of the spread statistics: IQR, Variance, Standard Deviation?

Numerical summaries of distribution:
   Five-number summary: [min, 1st Q, median, 3rd Q, max]
   Mean and standard deviation

Choosing a summary:
   Symmetric distribution, free of outliers: mean – std. dev.
Skewed distribution or one with outliers: 5-number or median – IQR

*How would you summarize the height distribution in this course (both sections)?*  
*How would you summarize the distribution of expected reading?*

**Boxplots:** Graphical displays of the five number summary
   Modified boxplots – 1.5*IQR “outliers” highlighted

*In class exercise: 1.61 on P. 60 – 61 M&M*  
*Data available in P:\Data\MATH\StatM&M5/Appendix\RUNNERS.MTP*

**Final Request for the week:** Take last two minutes of class to write down the ‘muddiest point of the week.’ What was nebulous or unclear to you? What do you not ‘get’ here in the first week of statistics?