In observational studies, it can be difficult to determine if causation exists.

If one were to experimentally change the native range of wild bunnies, and then saw a shift in ongoing military conflicts…

Quantitative Thinking in the Life Sciences

October 17th – Linking probability, mathematical functions and data
Part 2

Today

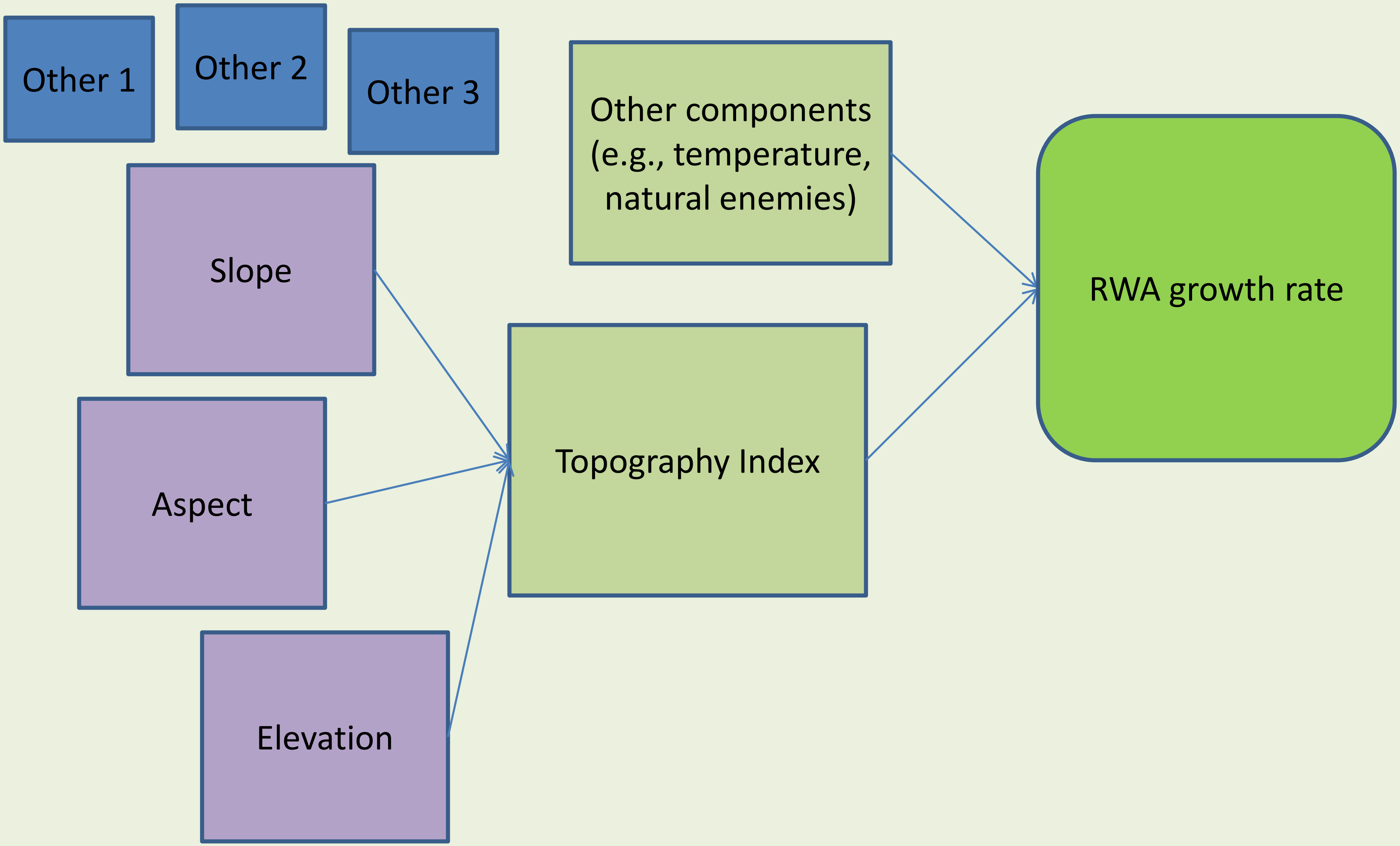
- Concept maps – Data distributions
- Simple mathematical relationships and probability
- Assignment # A
- More R fun!
  - R code questions?
  - Looking at snail vectors!

Housekeeping

- No homework turned in today
- New Homework # 6 is due on October 24^{th}
  - New assignment # 6 is part of old assignment # 4
    - Distributions and variability for your system's factors/components/variables
      - Distributions and variability estimates
  - I don't want to see anything about the relationships between factors (e.g., how x affects y) – unless I specifically e-mail you (e.g., Ali)

Data Distributions

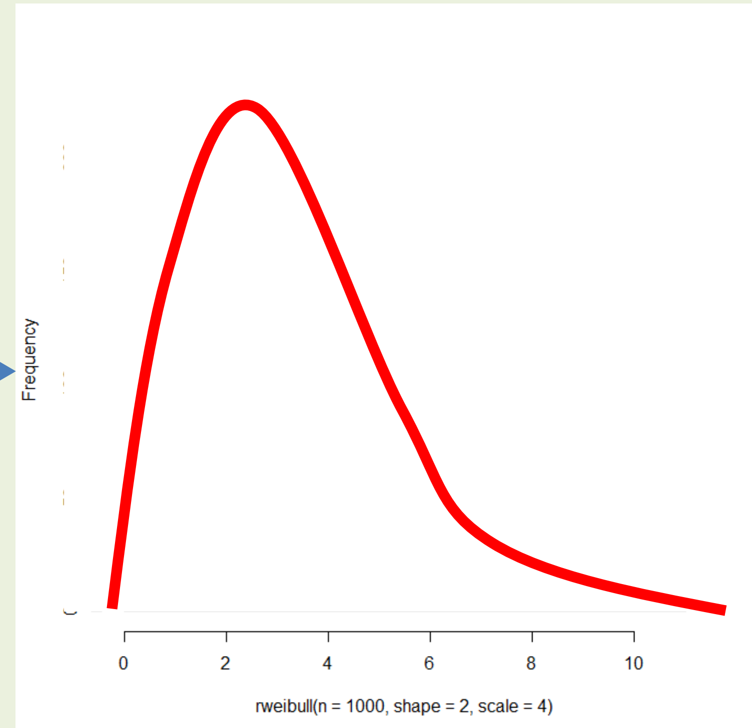
Russian wheat aphid (RWA) spatial growth rate model - concept map

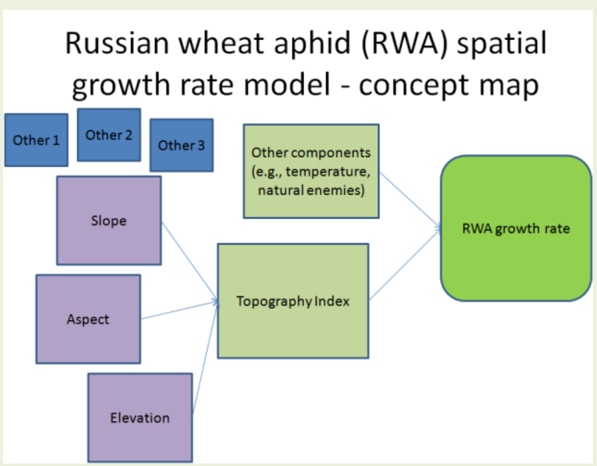


Slope component

Slope

- Slope in my wheat fields has limited variability
- Most of the fields are expected to have a slight slope with some brief sections with higher slope values

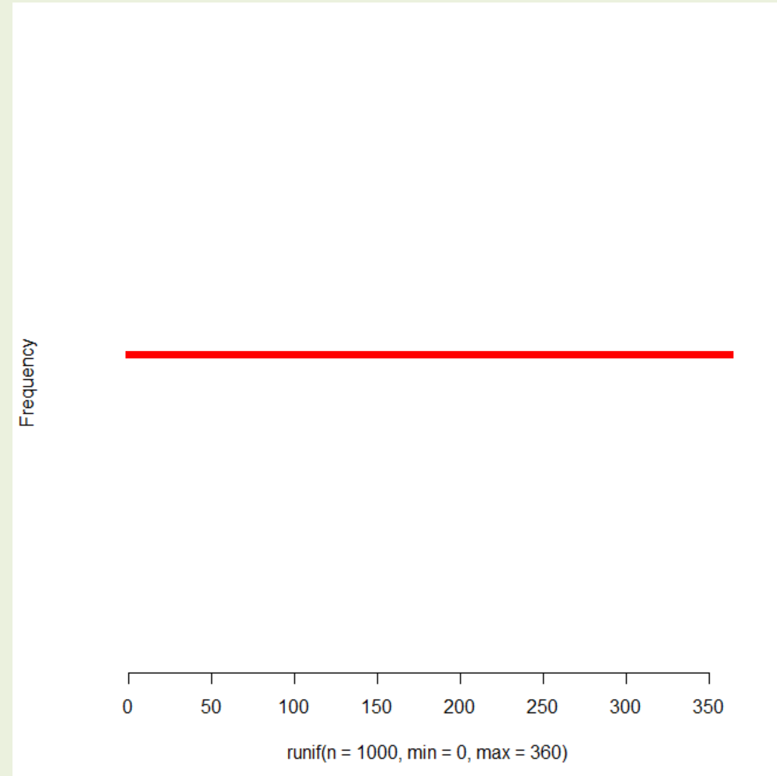


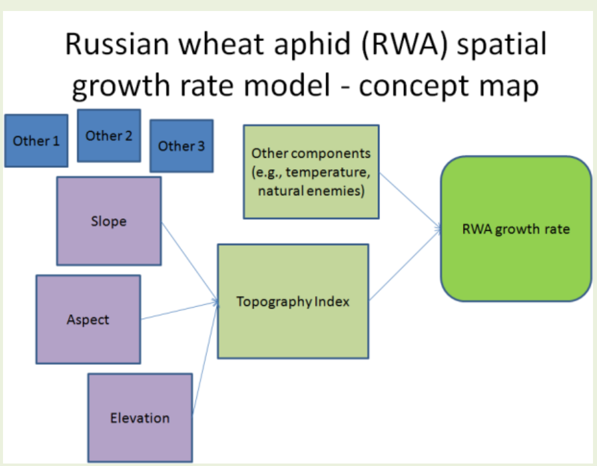


Aspect component

Aspect

- Given a slope value, I expect aspect to be uniformly distributed from 0 to 360 degrees
- That is, each plot has an equal probability of facing each direction

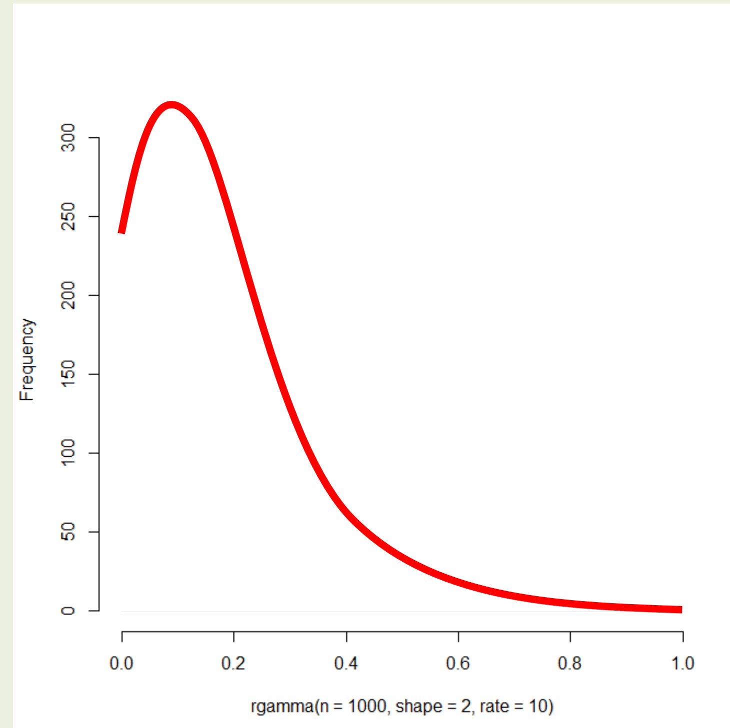


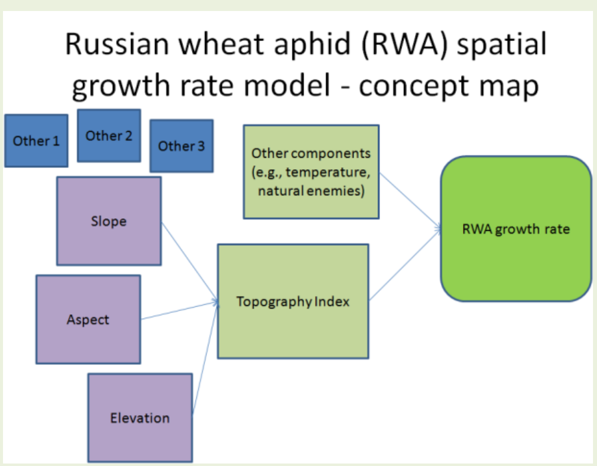


Relative Elevation component

Relative Elevation

- most of the field is relatively flat but with a couple of terraces and a small hill
- That is, most plots will be at relatively low elevations with some exceptions

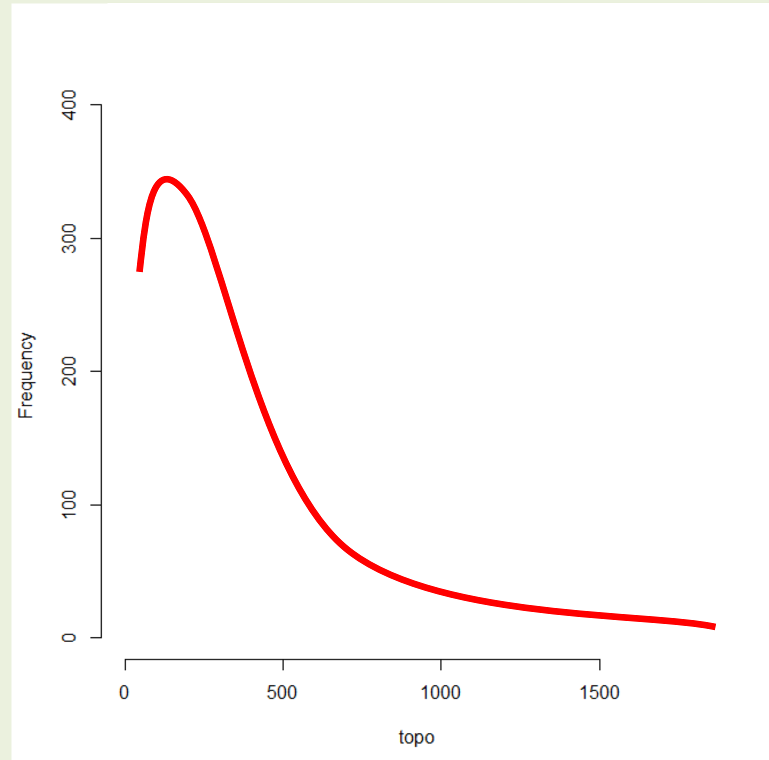




Topography Index

Topography Index

- I don't know exactly, but given that I have right skewed and uniform distributions as inputs, I likely will have a right skewed topography index



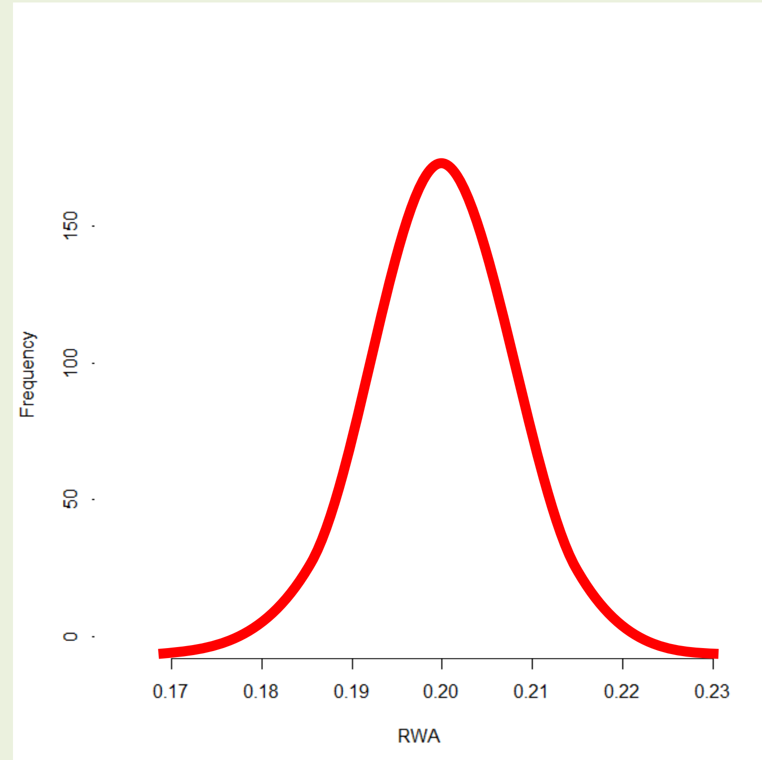
Russian wheat aphid (RWA) spatial growth rate model - concept map

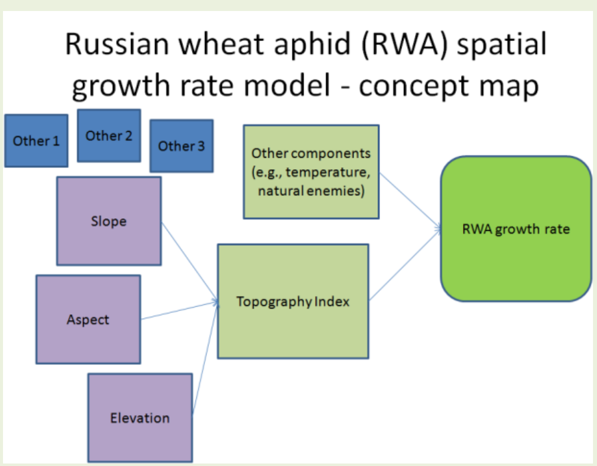
BTW, right skewed means that the tail is on the right!

RWA growth rate

RWA growth rate

- I imagine after all the components are factored in, growth rate data will be fairly normal but may be slightly right skewed

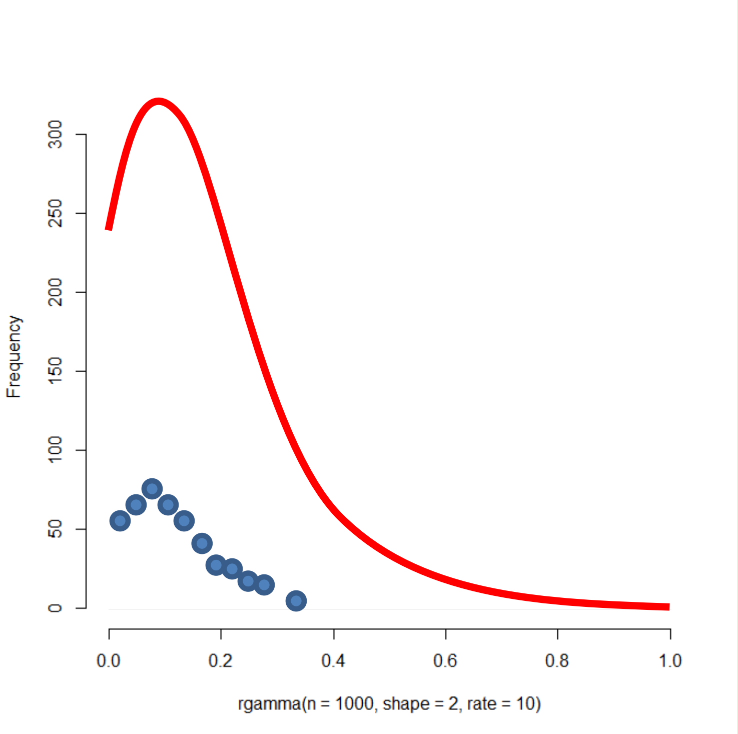




Why is data distribution important?

Relative Elevation

- Most of the field is relatively flat but with a couple of terraces and a small hill
- That is, most plots will be at relatively low elevations with some exceptions



If we sampled randomly without stratifying, we could end up only sampling a very narrow range of relative elevation values.

It is hard to tell there if there is an effect of high relative elevation if no high relative elevation data were obtained.

For the homework

e.g.,

- What distribution of P data do you expect to be able to collect from the effluent?
- What will your blueberry yield data look like?

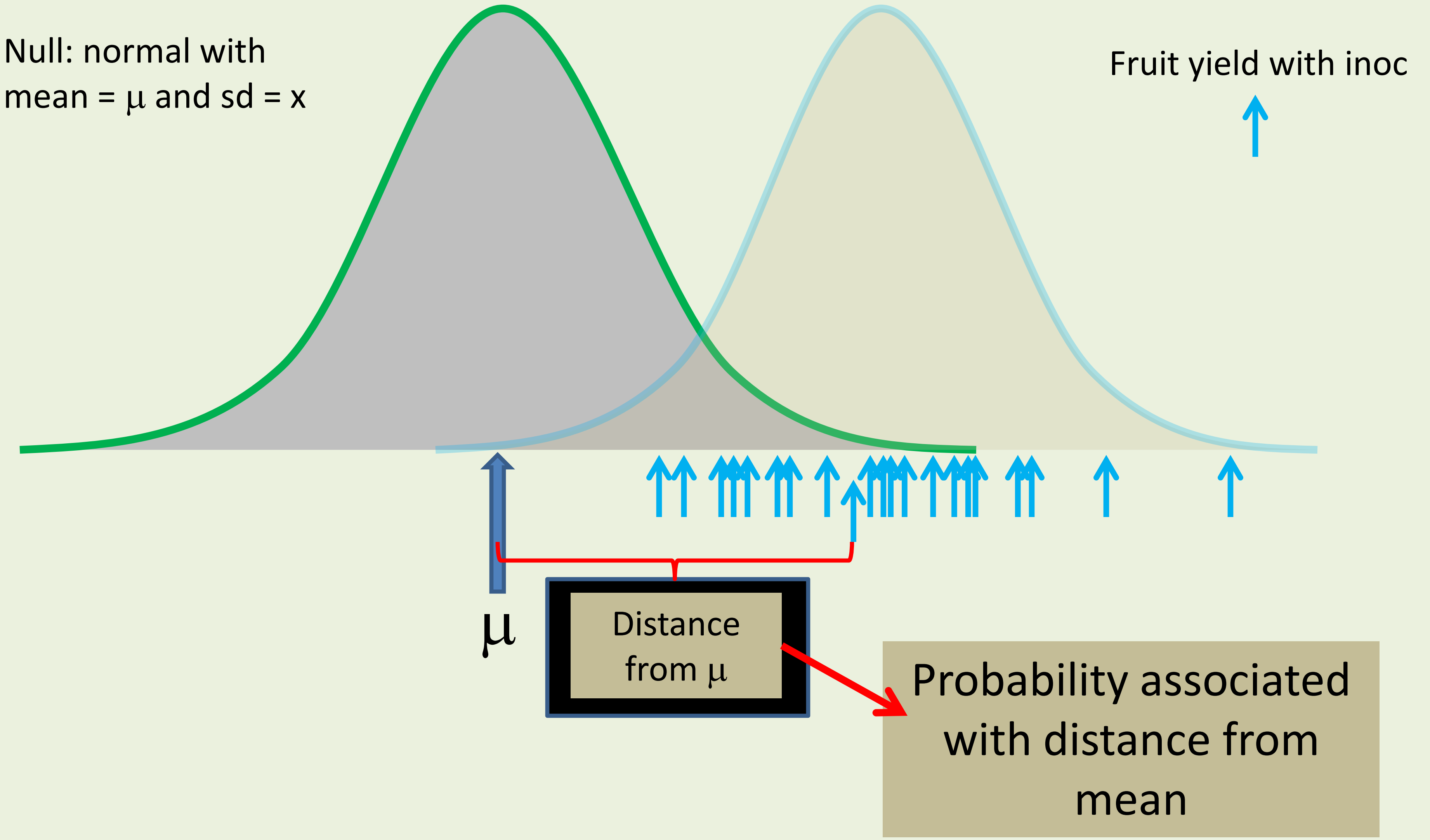
I do not want to know how temperature will effect growth.

I want to know what the temperature data will look like.

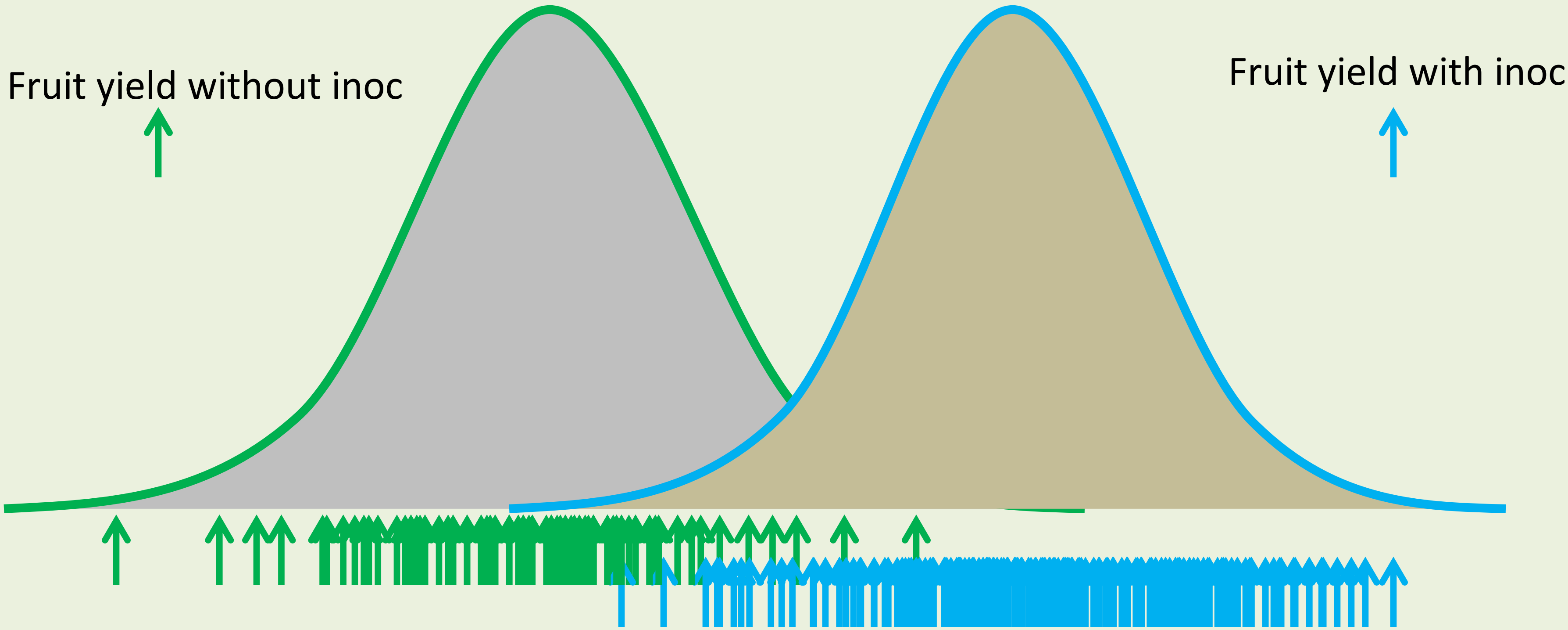
- X-axis should have the component (e.g., mm storm H_20)
- Y-axis should be relative frequency of observing the data

Onward!
Variation in your data

Observed data connected to the p-value

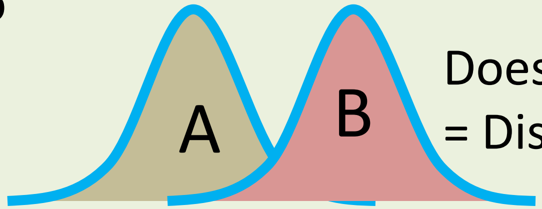


Do the distributions of A and B differ?



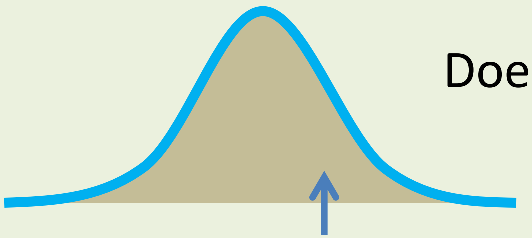
t-test will allow us to test

Test a null hypothesis that the means of two normally distributed populations are equal



Does Distribution A = Distribution B?

Test that a population have a mean value (specified as your null hypothesis).



Does $\mu = 5$?

Paired or repeated measures test (collect data from something twice and see if the data differ).

Miticide Trial Data

Mites/Plant	Before	After
Corn plot 1	0.500	22.967
Corn plot 2	10.657	29.364
Corn plot 3	43.469	15.972
Corn plot 4	7.045	7.683
Corn plot 5	9.626	10.089
Corn plot 6	18.534	14.059
Corn plot 7	34.237	23.093
Corn plot 8	38.291	28.351
Corn plot 9	11.959	4.898
Corn plot 10	1.582	13.964

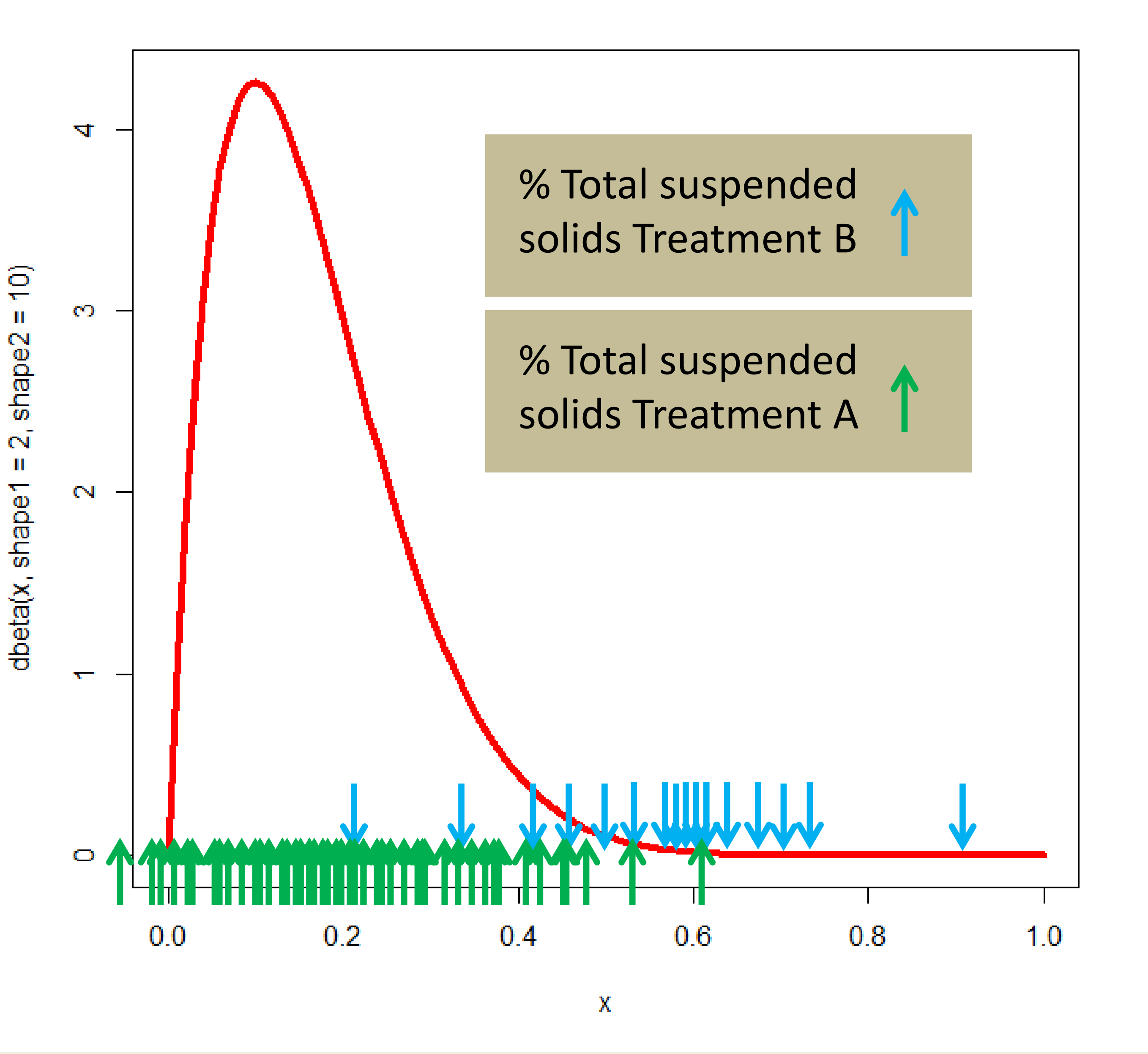
Does Treatment A change the population?

Distributions matter!

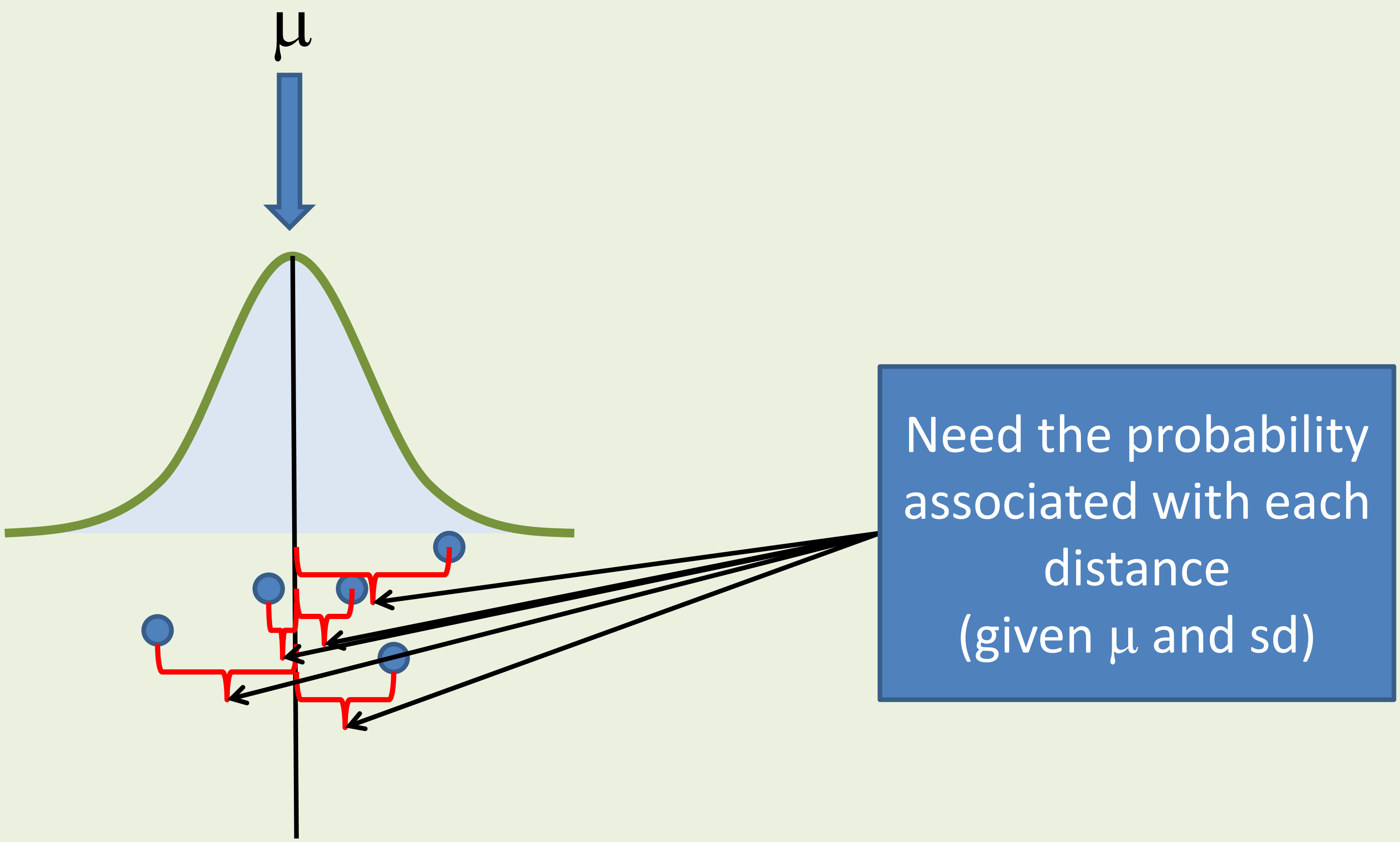
Beta Distribution

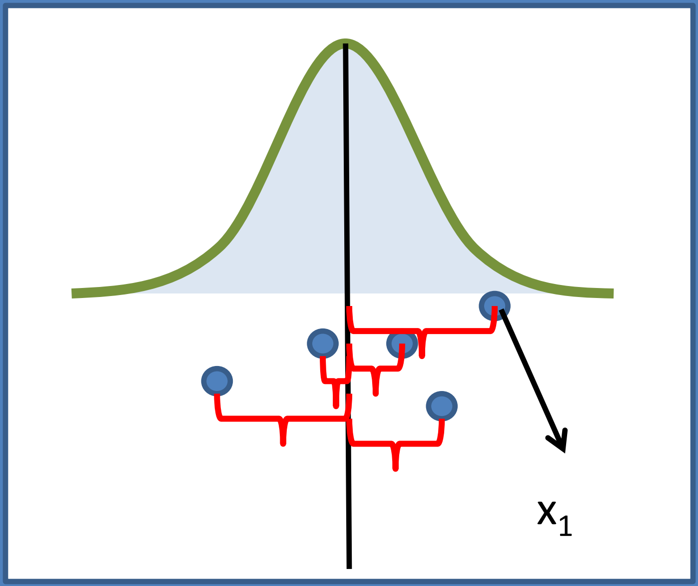
Shape 1 = 2

Shape 2 = 10



Developing a test statistic with a normal distribution

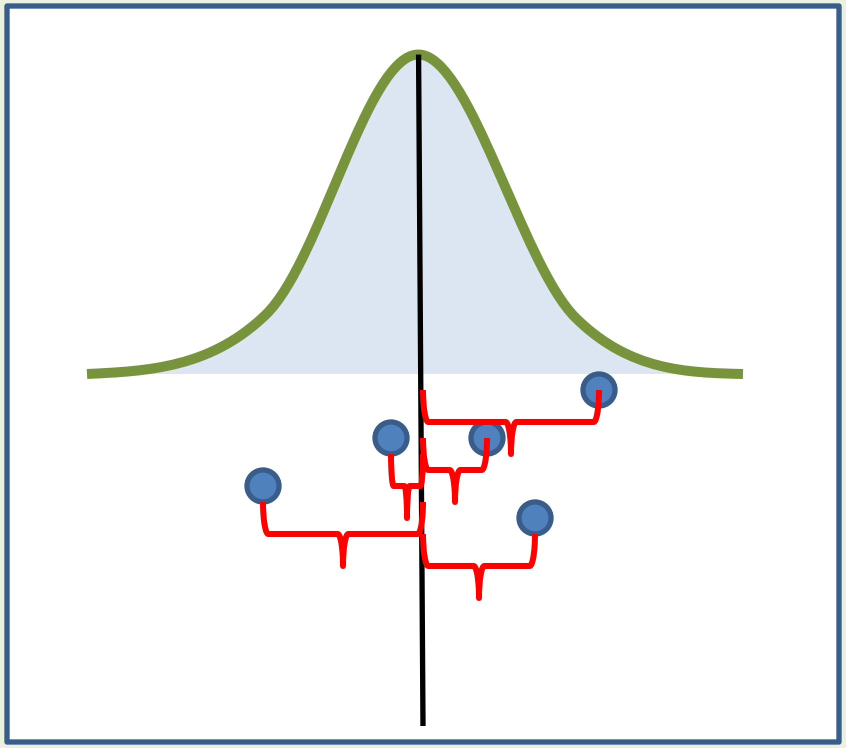


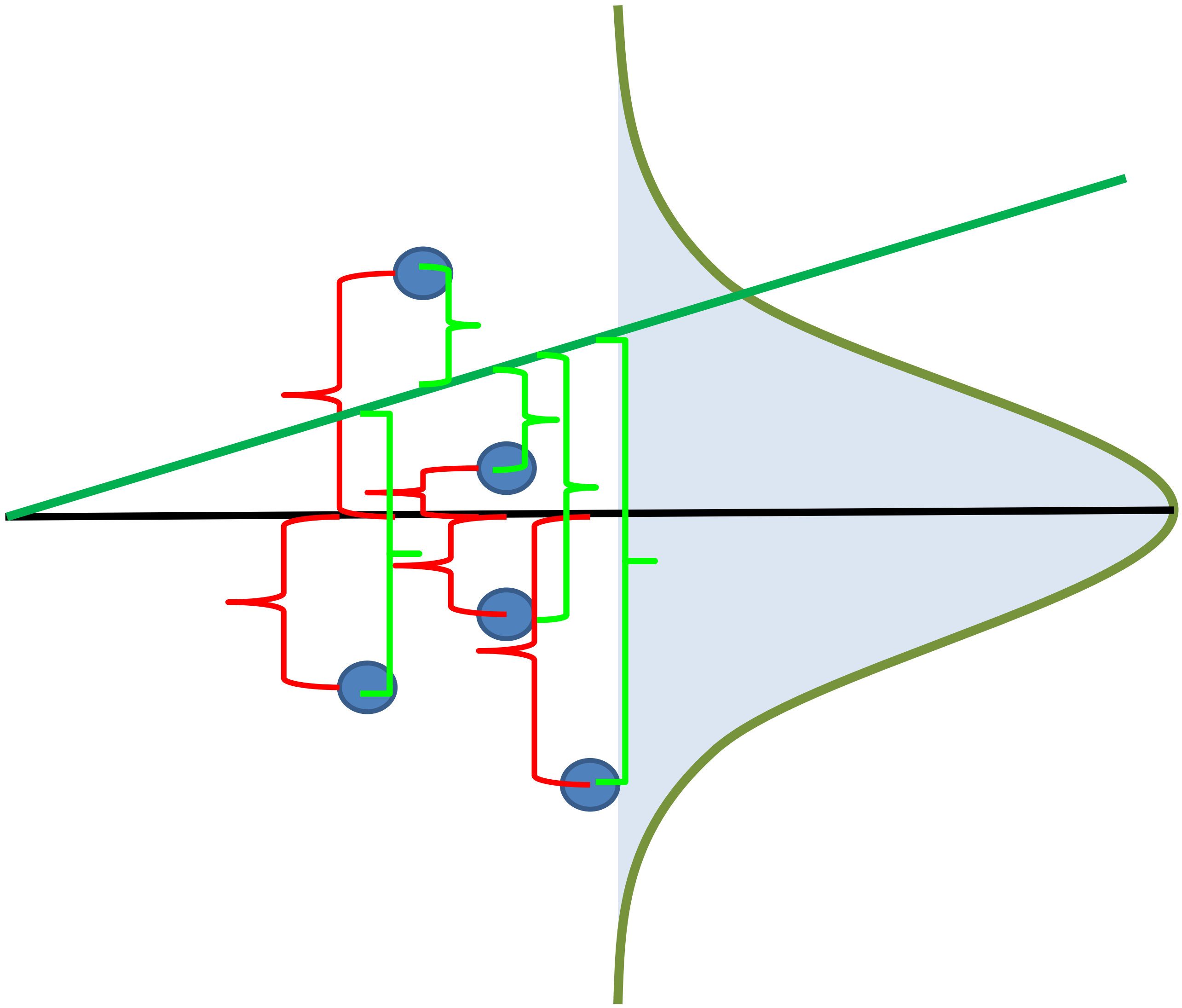


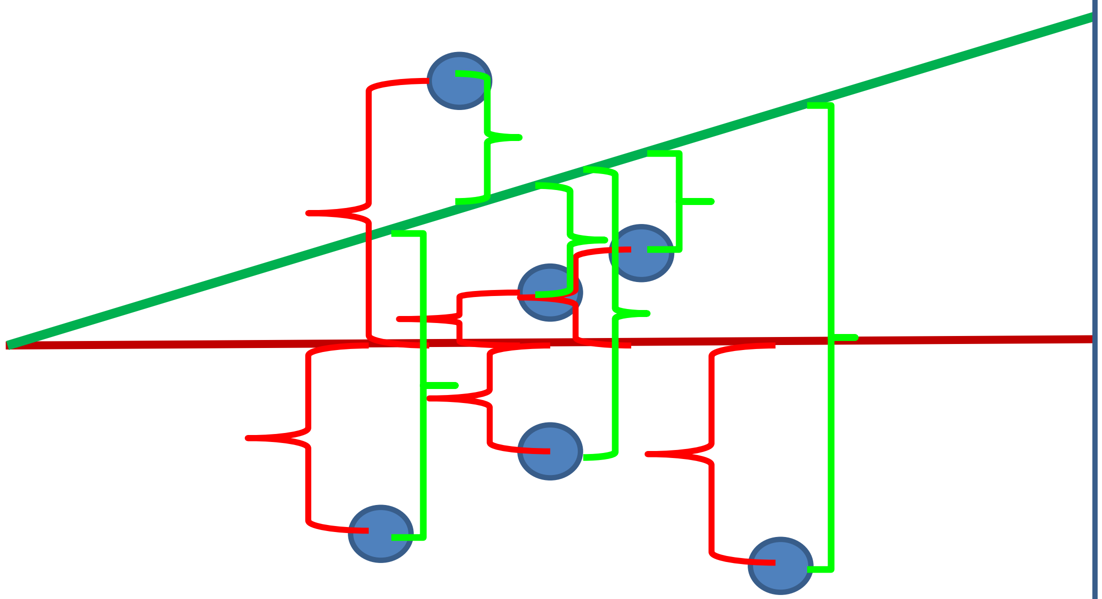
$x - \mu$ = Distance

Allows us to quantify the probability of x's occurrence

Time check!

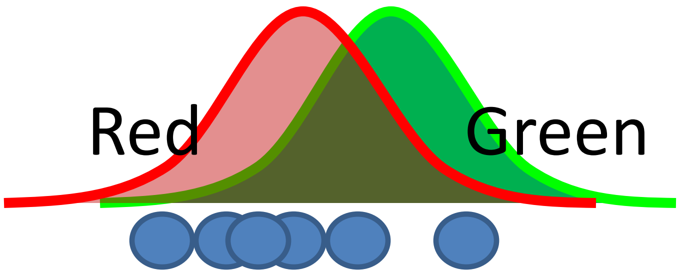






Distances with slope = 0

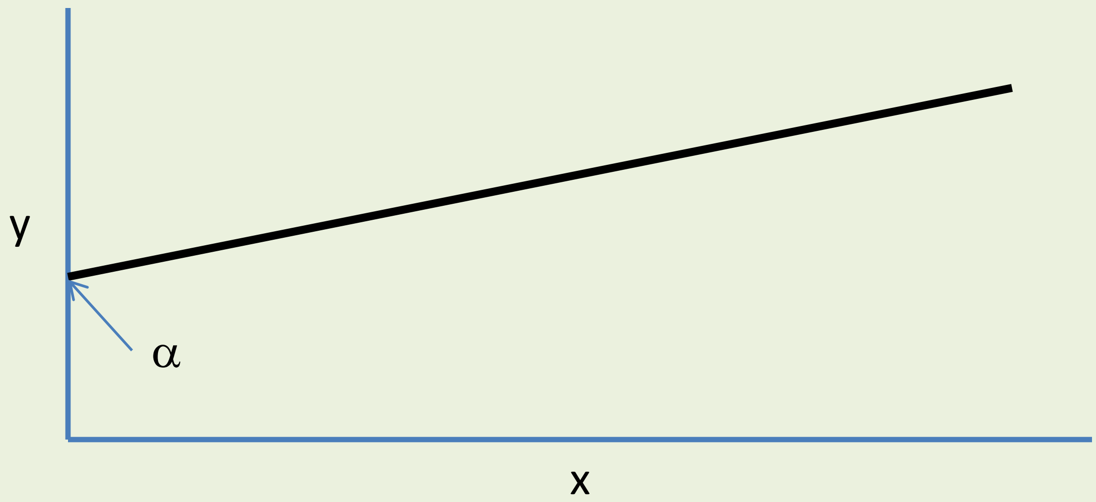
Distances where slope has a value



Example in R!

– Cow Weight in Dropbox!

Linear model: $y = \alpha + \beta_1 * x$



$\beta_1 = \Delta x / \Delta y$

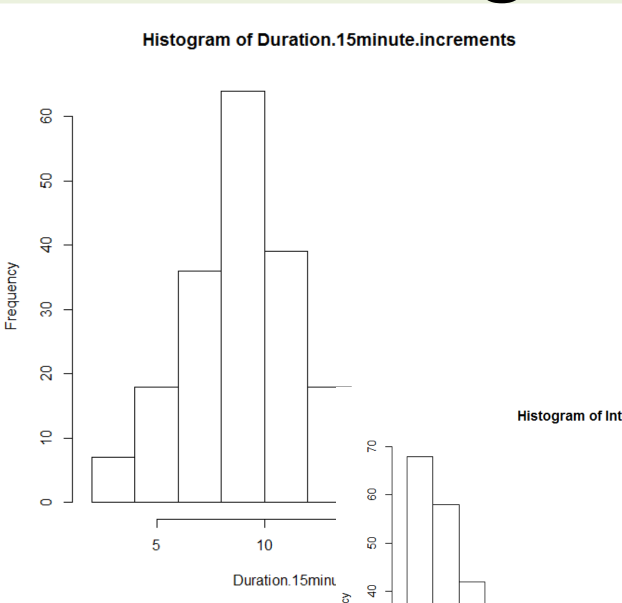
Assignment # A

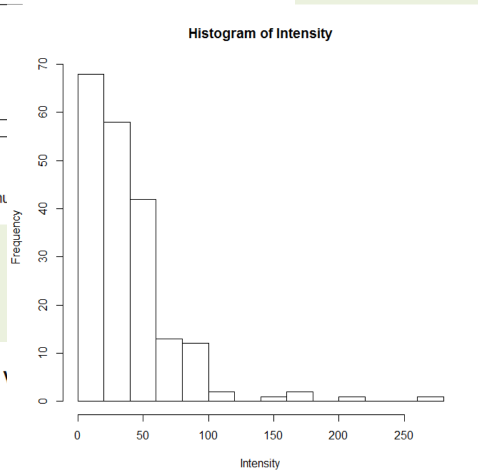
- Assignment # A is due on October 24th
- Worth 50 points
- Concept Map Distributions
  - Describe data distributions for your components (not relationships between the components)
  - Write up in paragraph form plus an introduction and figures
- Part 2: Chapter 7 R code found on my website
  - Distribution exercises and examples for use in future simulation work will be in this chapter

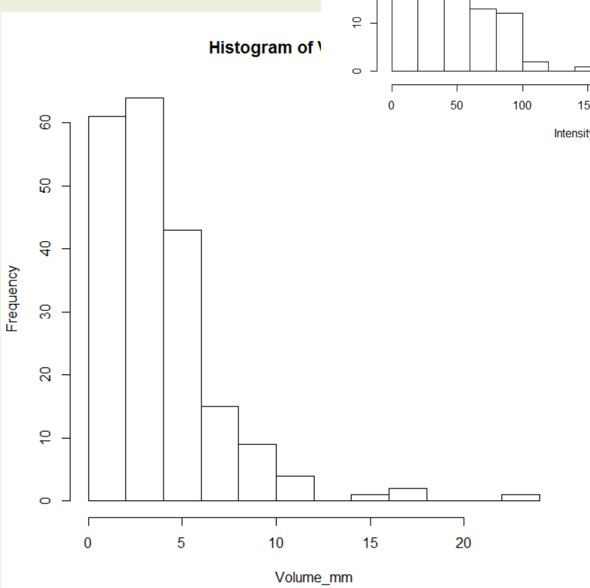
Assignment # B

- Assignment # B is due on November 1st
- Worth 50 points
- Part 1: Simulation
  - Using the provided functions for distributions, take a first pass at simulating data for each of your components where you will be taking data. Assume that data will be measured perfectly (no measurement error).
  - Write up in manuscript form for a few of the components. That is, introduce the system (you can self-plagiarize but make it clean), describe how you will sample (or already sampled) components (Methods section), describe your simulation inputs, include output plots. Discuss in brief.
- Part 2: Chapter 8 R code – not posted!

Testing a Bioretention systems: Total Suspended Solids





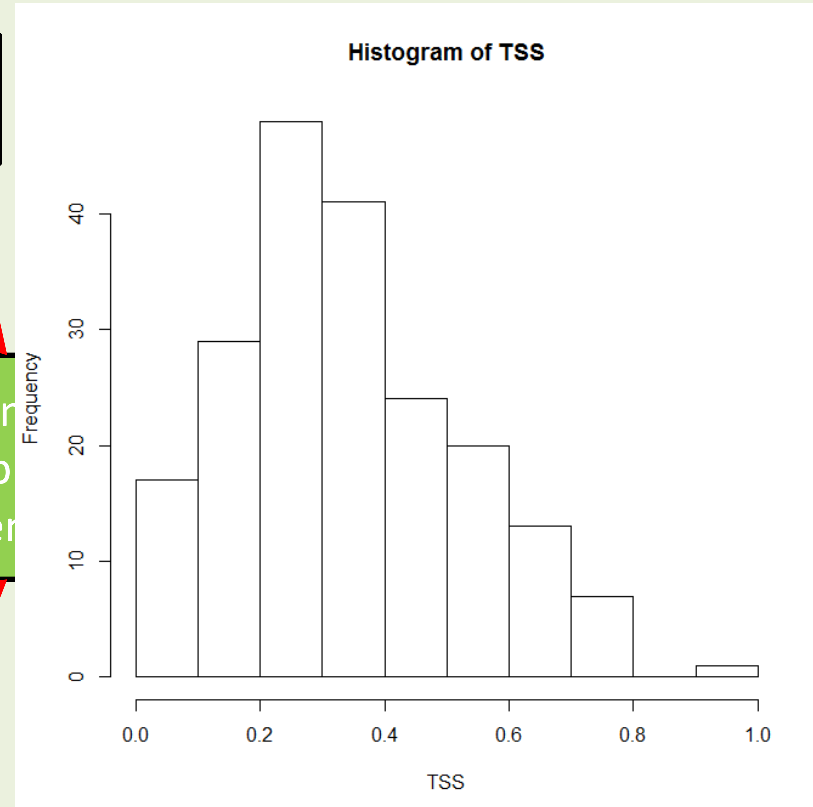


Poisson, most events around 2.5 hours

Duration

Volume

log normal, 3mm mean, 2 mm standard deviation



Liner Strip

Cell vegetation