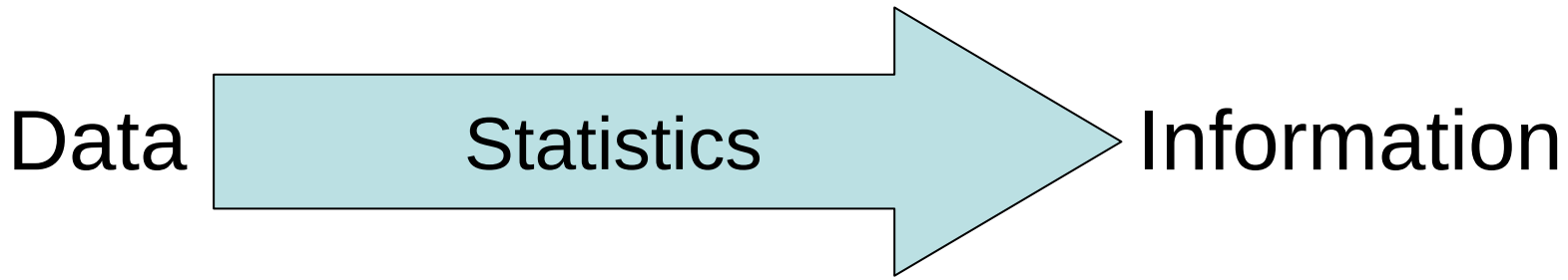
Statistical Toolset for Maximizing Information From Function Point Data

Dwayne Pepper
Intel Corporation
5000 West Chandler Boulevard
Chandler, Arizona 85226
United States



Statistics: The science of gathering, summarizing, presenting and making inferences from data.

This presentation will add a few tools to your problem solving toolbox.

- "Applications of Statistical Thinking…" was presented in the Spring 2006 FSS. This paper motivates the need for statistical thinking in organizations that use functional sizing metrics. This abstract is for a follow-up paper that leads participants through the core expectations for statistically valid use of function point data. Topics include avoiding statistical malpractice in summary presentations, the essentials of blocking and using correct sample sizes for inferences and estimations.

- Presented is development of a statistical toolbox that has been widely distributed and used across Intel's IT organization. The presentation starts with a definition of statistical malpractice and leads participants through how to avoid it. This proves essential to deliver statistically valid summary presentations. Project size estimation is used to introduce the statistical inference notion of confidence intervals. This leads to a classical treatment of statistically valid sample size determination. The importance of blocking is introduced through function point based metrics by platform, development types, etc.

Why we need statistics . . .

- Introduction
  - Something's new at Intel
- Problems encountered
  - It's scary…..and can cause problems
- Solutions
  - But we won
- Opportunities
  - This is widely fungible
- Critical statistics toolbox
  - Statistical malpractice
  - Blocking
  - Inferences
  - Sample sizes

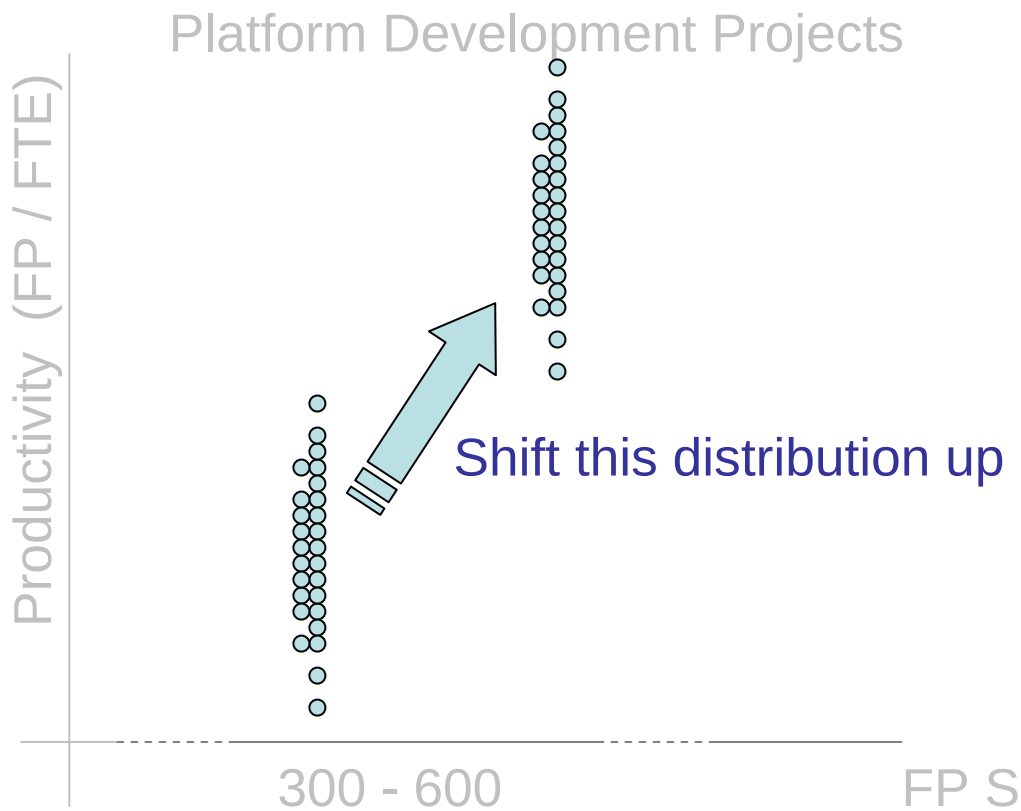
Introduction

- Function Point-based metrics are:
  - Relatively new to Intel
  - Generally accepted as essential to continuous improvement within our Information Technology (IT) group
    - Essential to normalize **Productivity**, **Cost**, **Duration** and **Defect Density**
  - Intended to provide high-level drive or motivation for improvement
    - Expected to follow a path toward more tactical usage over time

Problems Encountered

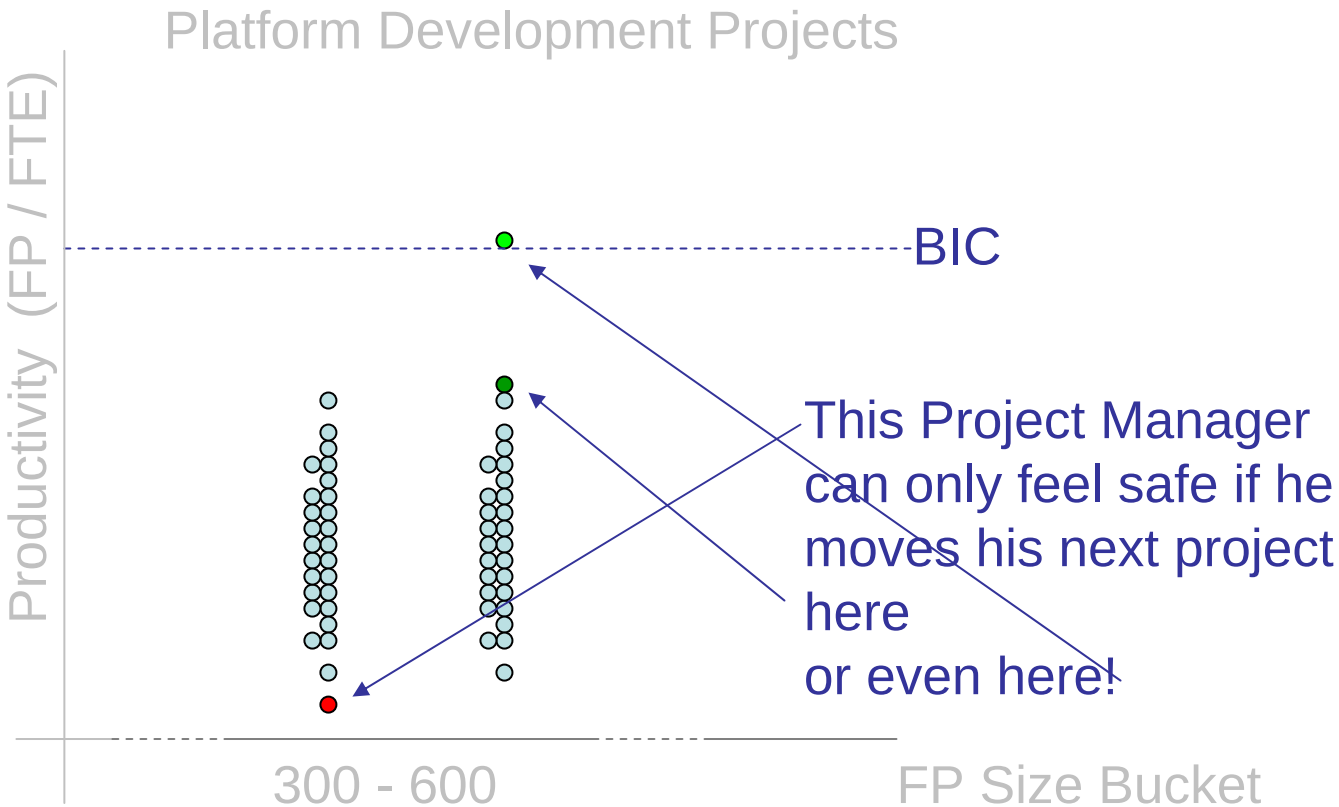
- Intended to provide high-level drive or motivation for improvement.
- Neither our Developers – nor anyone else at Intel – are accustomed to individual productivity metrics.
- Intel Culture immediately gravitated toward individual performance management.

Our Goal



This presentation contains 0% real FP – based data.

Initial Result

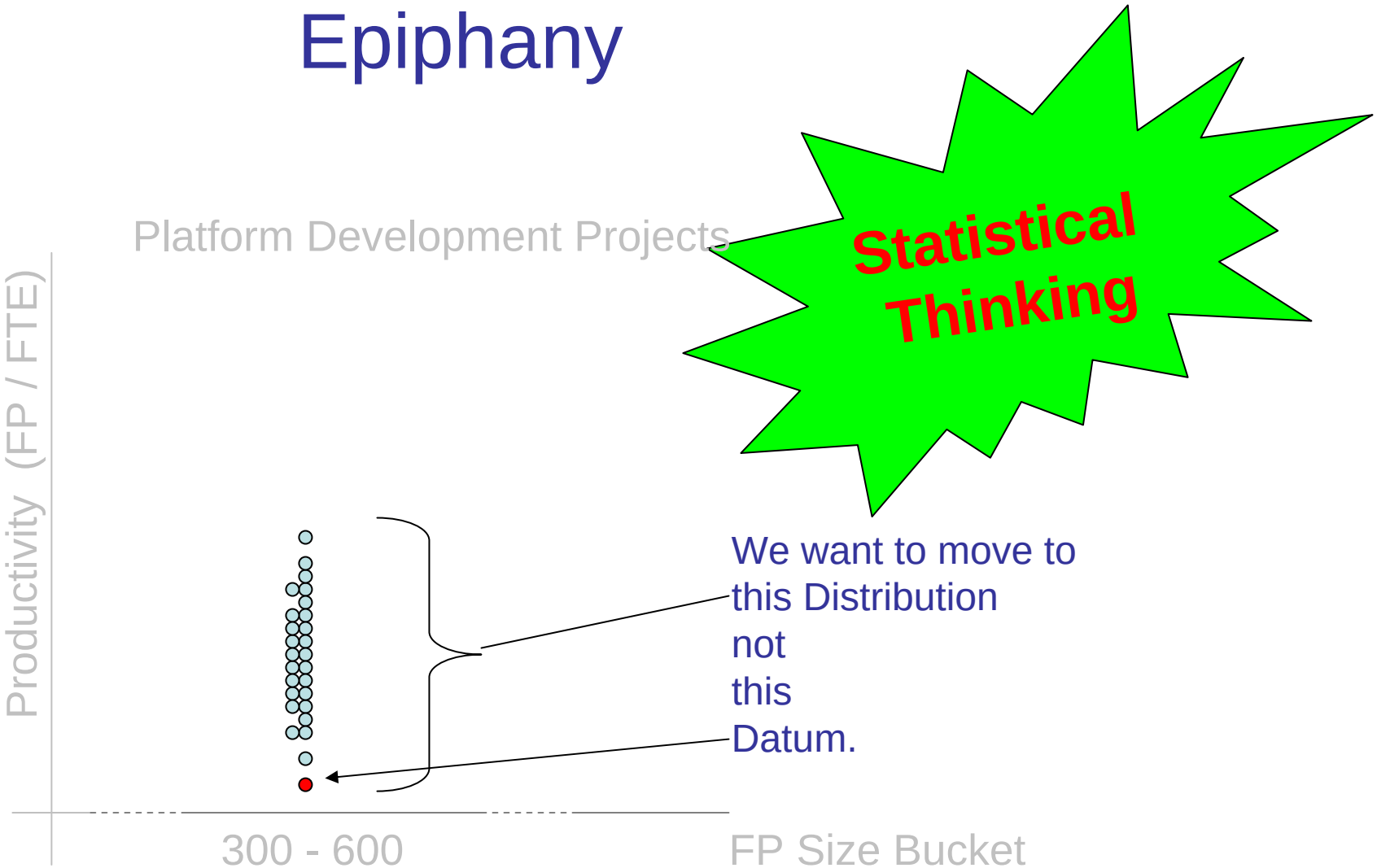


Slowest gazelle management

Problems Encountered

- All projects being best-in-class for productivity is not possible.
- Likely self-destructive to pursue.
- Objective is to identify future productivity improvement opportunities – not to punish previous performance.

Epiphany


Platform Development Projects
Statistical Thinking
Productivity (FP / FTE)
We want to move to this Distribution not this Datum.
300 - 600
FP Size Bucket

Solution

- H. G. Wells: “Statistical thinking will one day be as necessary for effective citizenship as the ability to write."

- The philosophy in Statistical Thinking is widely used within Intel’s manufacturing environment. It provides an avenue for correct usage of our new metrics.

Solution: Statistical Thinking

- All work is done in a process…*and that process can be measured.*
- All processes have variation...*and that variation is reflected in its measurements*
- Reducing the variation will improve the process…*and we better be interested in improving our processes*.
- Appropriate action can not be taken until you understand your variation.

Codified by American Society for Quality c. 1996

Solution:
Statistical Thinking

- Evidence of Statistical Thinking must be shown in our presentations.
- Getting the correct message from our data depends on it.
- The key is “…understanding our variation.”

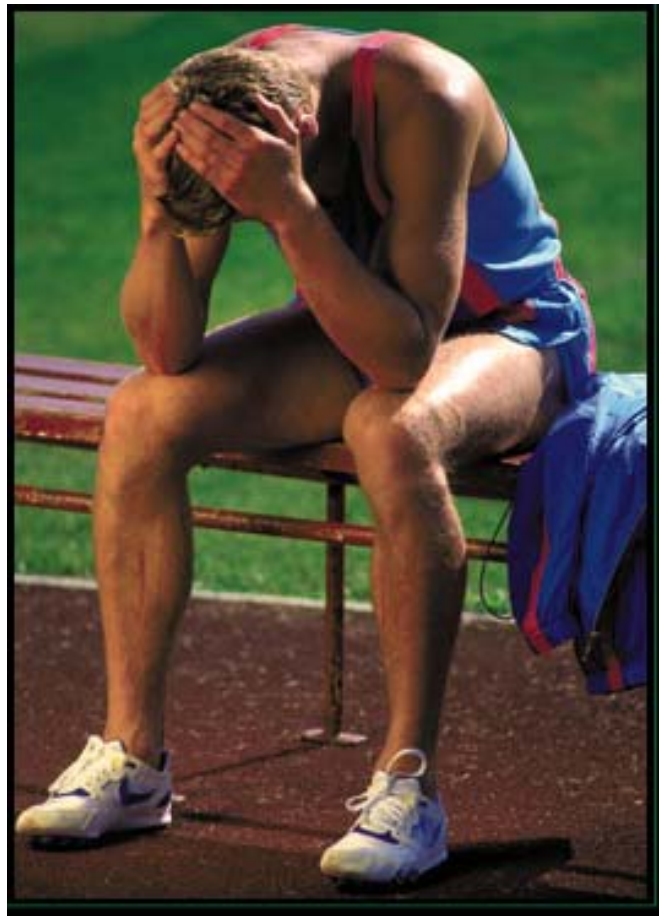
Avoiding Statistical Malpractice in our Analyses and Presentations

Definition: *Statistical Malpractice*

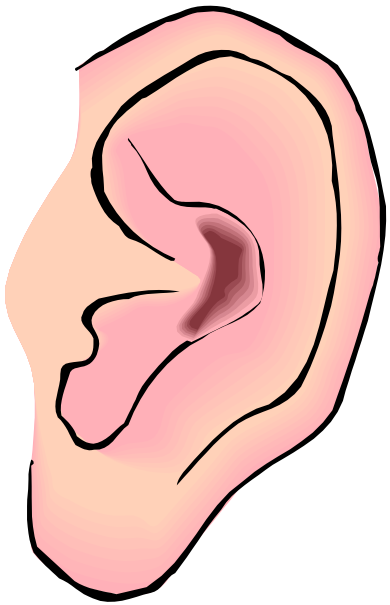
- The act of misleading, misinforming or confusing others by means of an incomplete or erroneous statistical analysis of data, whether due to intent or accident.

Definition: *Statistical Victim*

- One who as been misinformed, misled or confused by an incomplete or erroneous statistical analysis. Also know as *Victim.*

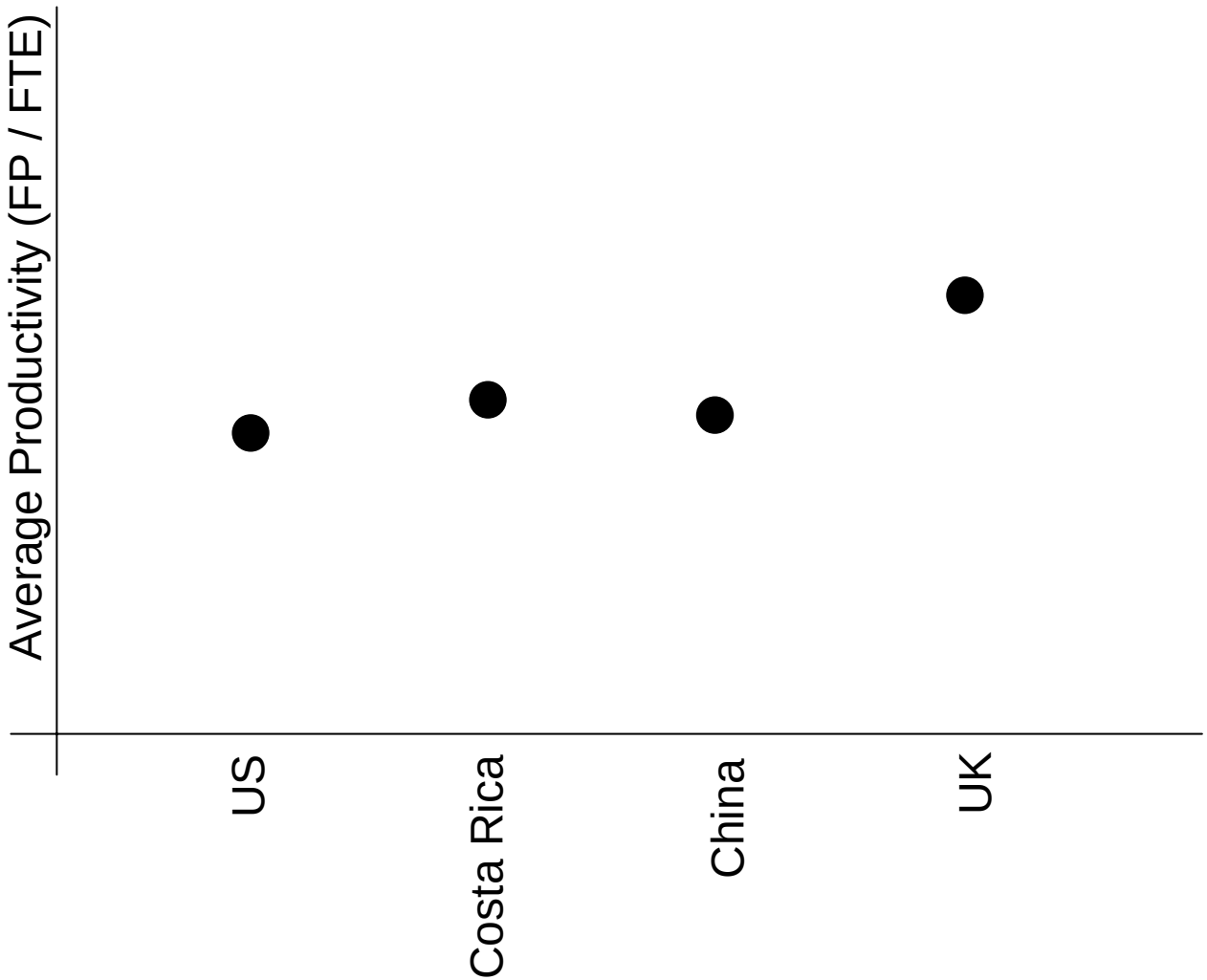


Common tools to avoid being a statistical victim:

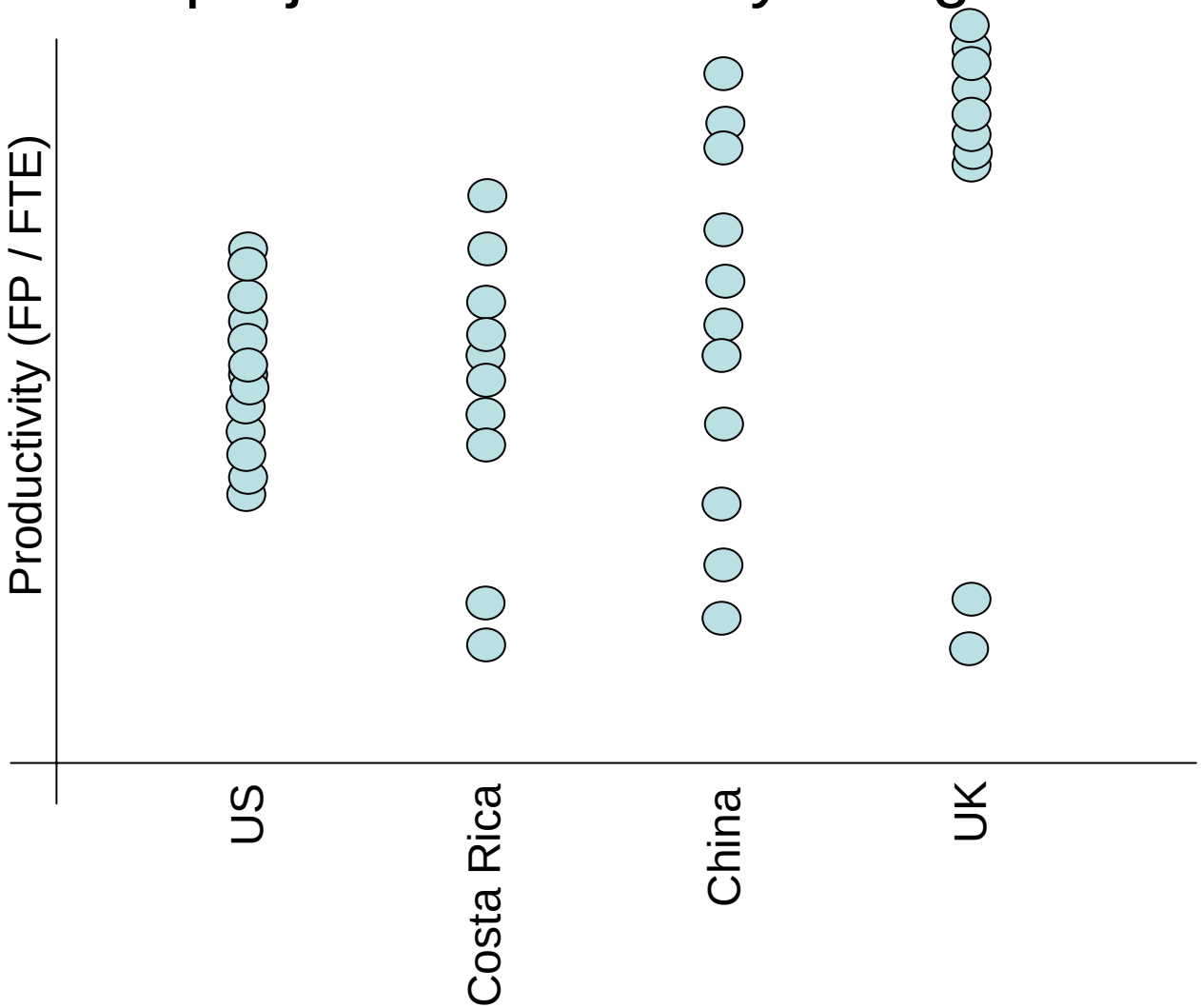


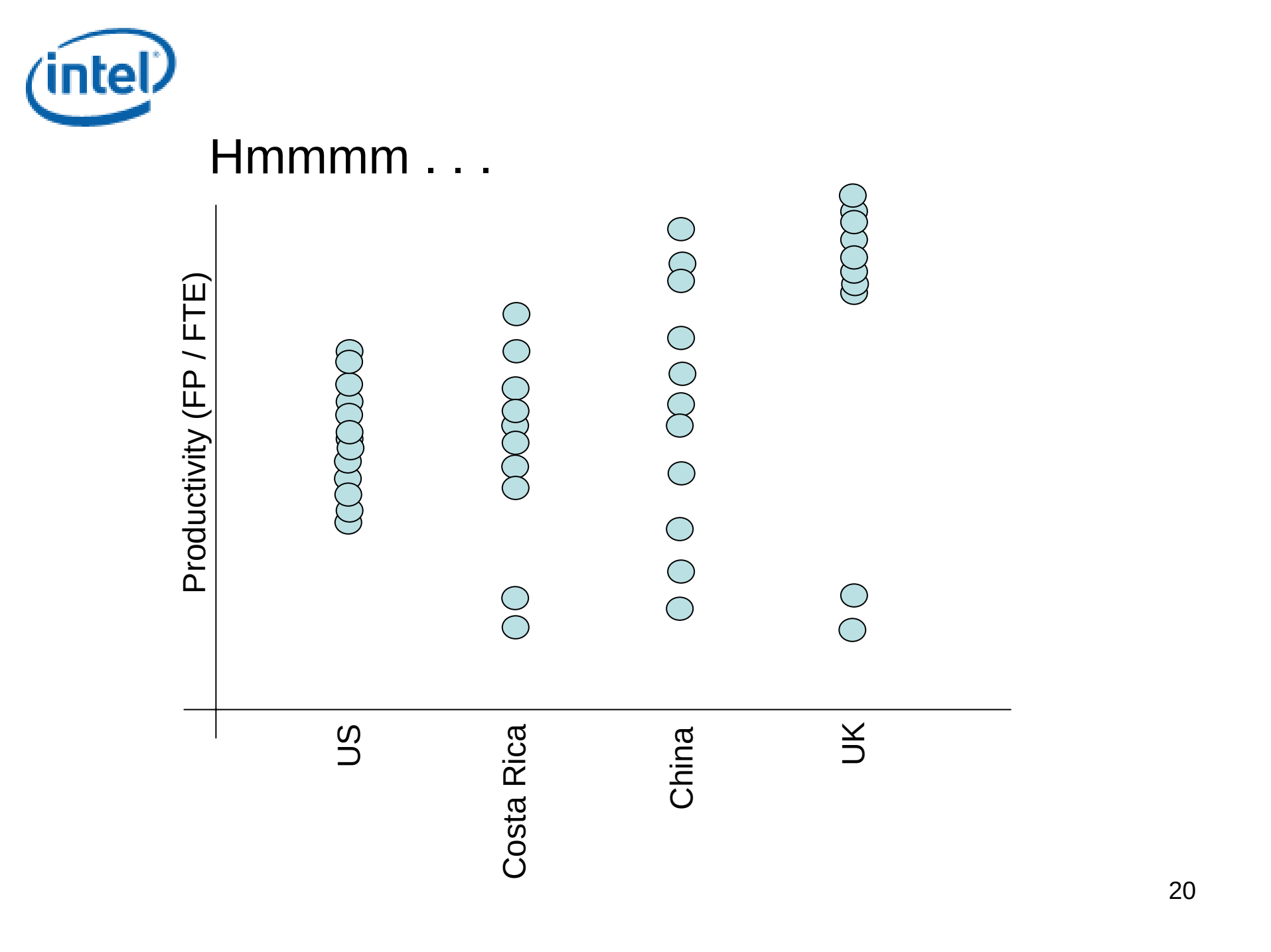
The Ear

Our project Productivity is highest in UK.



Our project Productivity is highest in UK.





Hmmmm . . .

Productivity (FP / FTE)

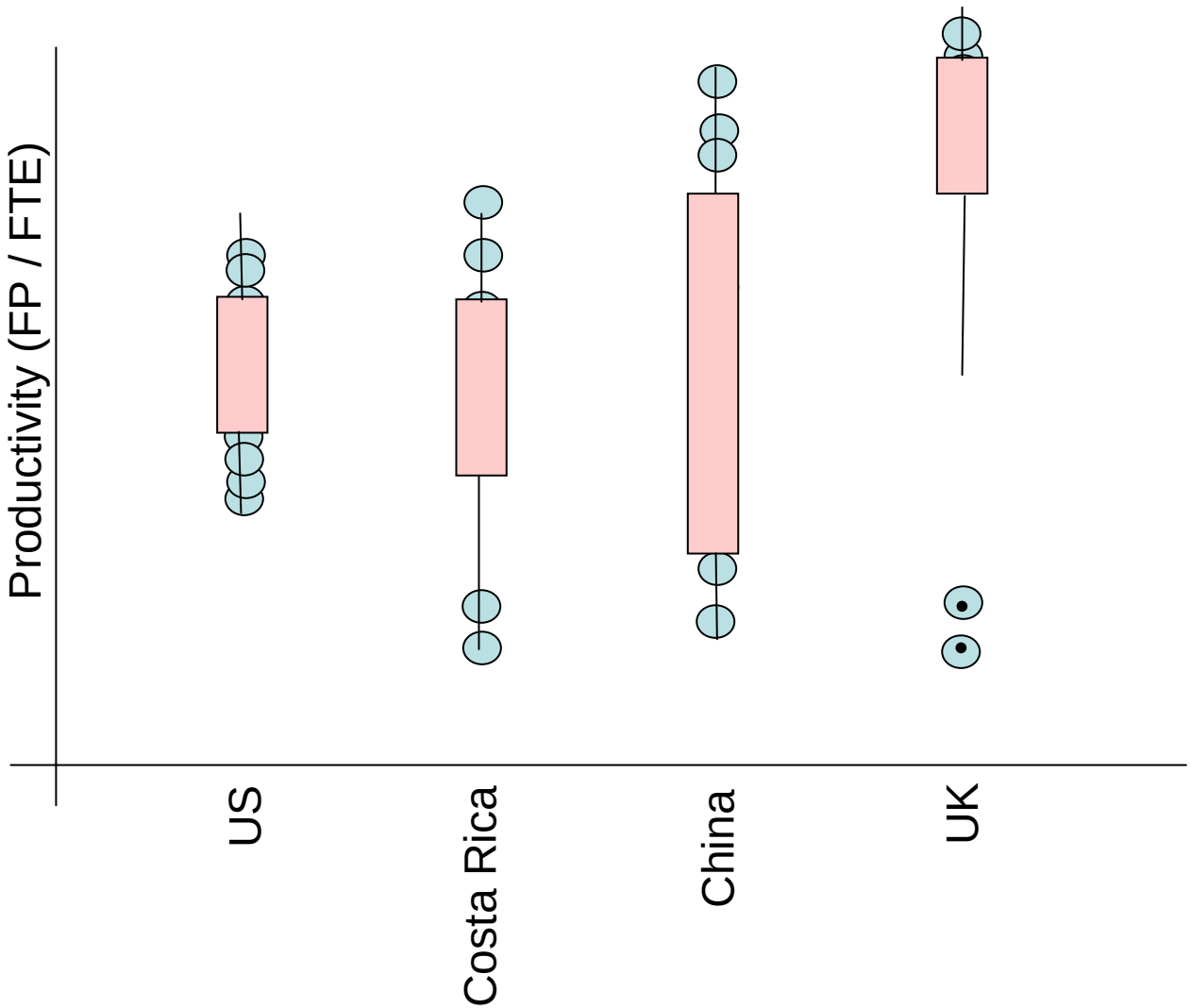
US

Costa Rica

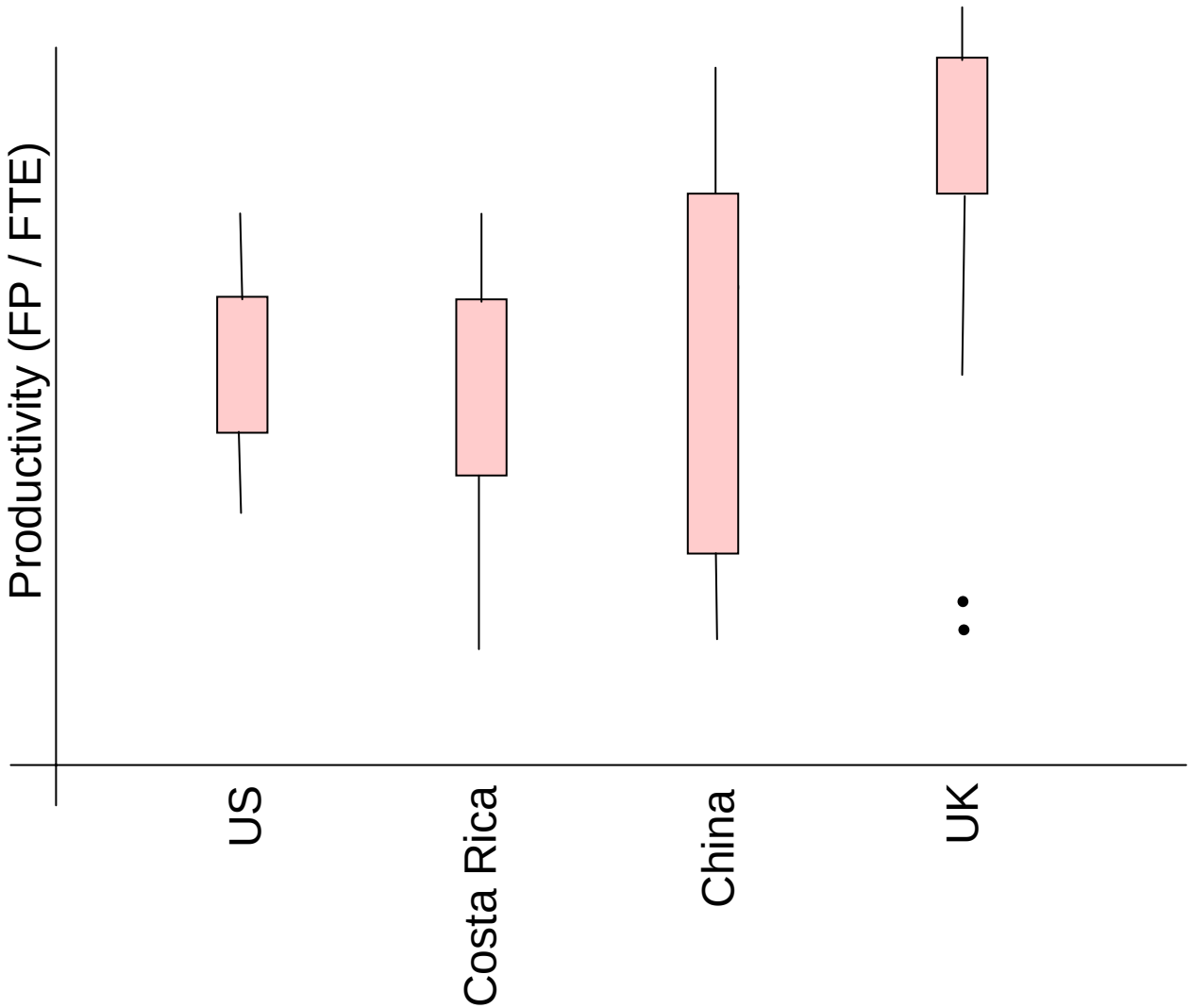
China

UK

Another representation - Boxplots



Another representation - Boxplots



Statistical Thinking

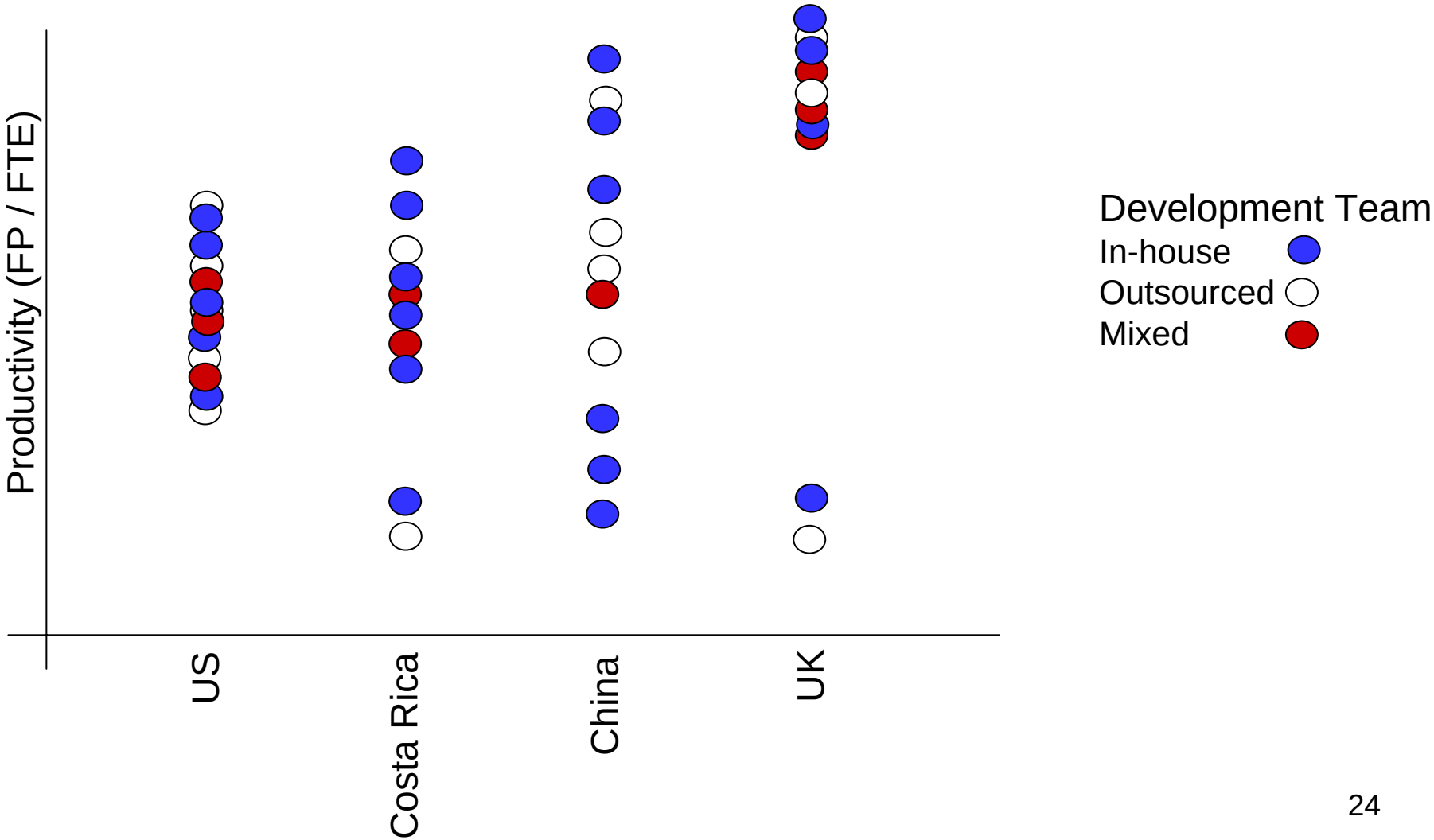
Appropriate action can not be taken until you understand your variation.

What do I understand about my variation?

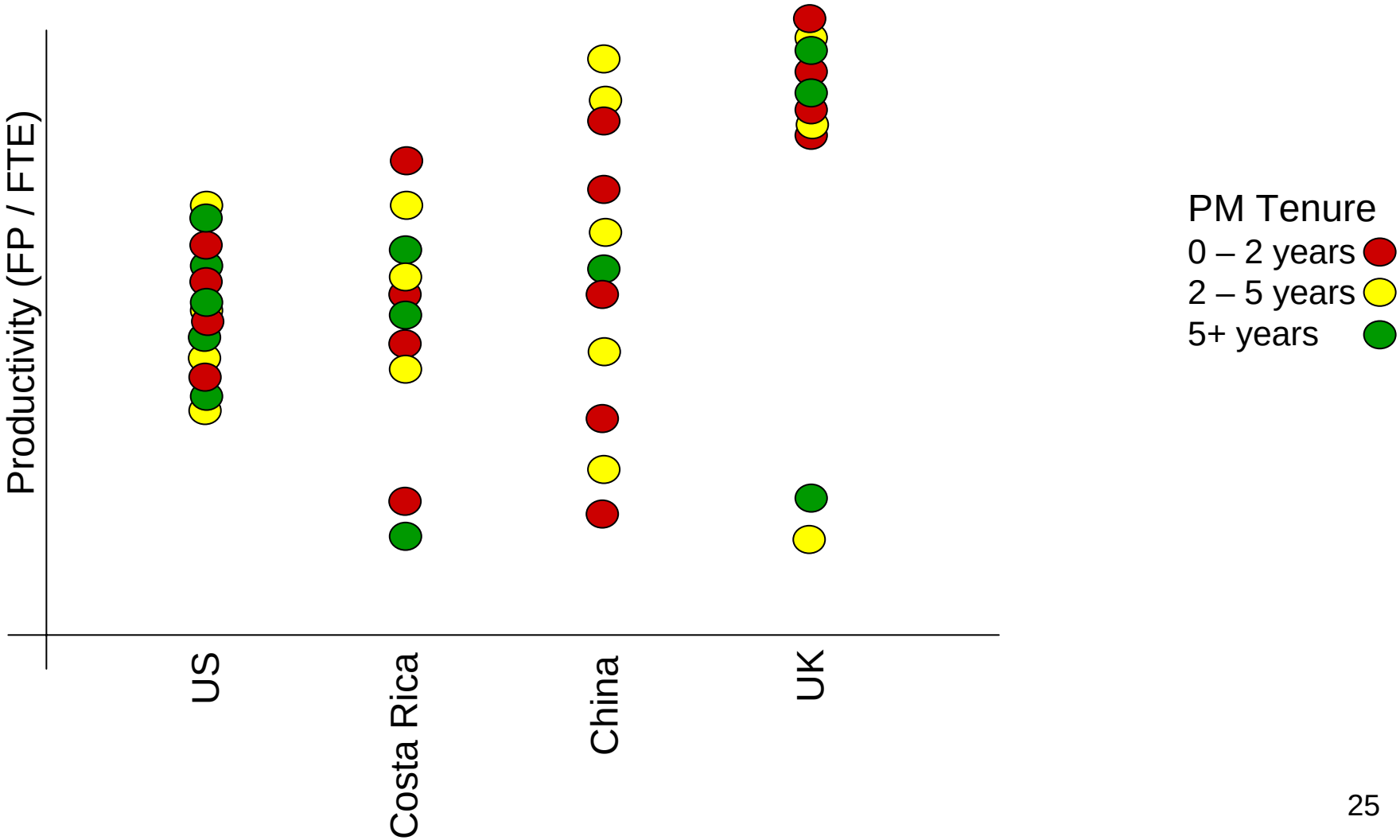
- Cannot understand variation based on a single point.
- Must identify and account for as many known sources of variation as possible
- Test for unknown sources of variation
- Categorize remaining variation as natural or special cause

Evidence of Statistical Thinking must be present in our presentation.

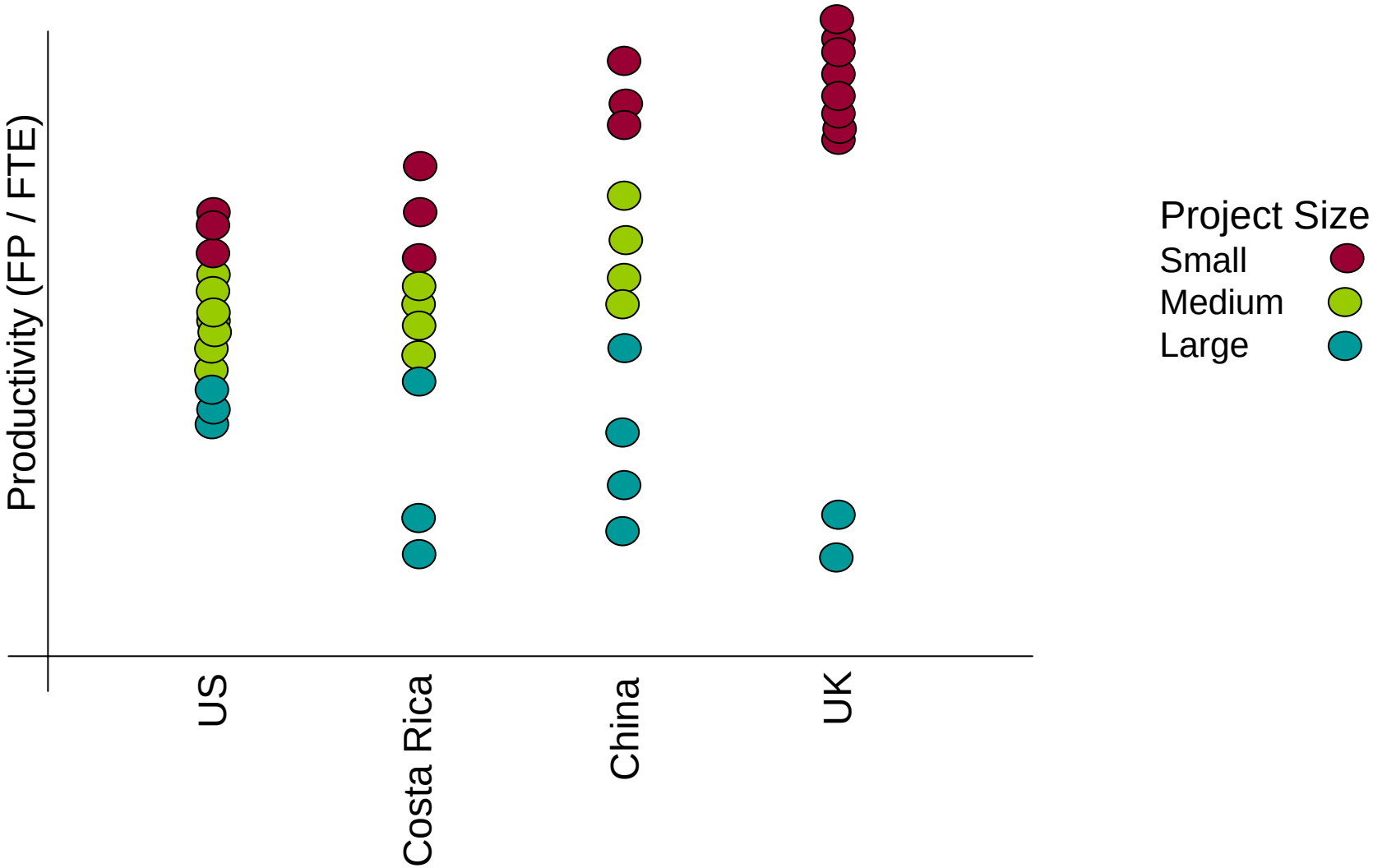
Understanding Variation, my way . . .



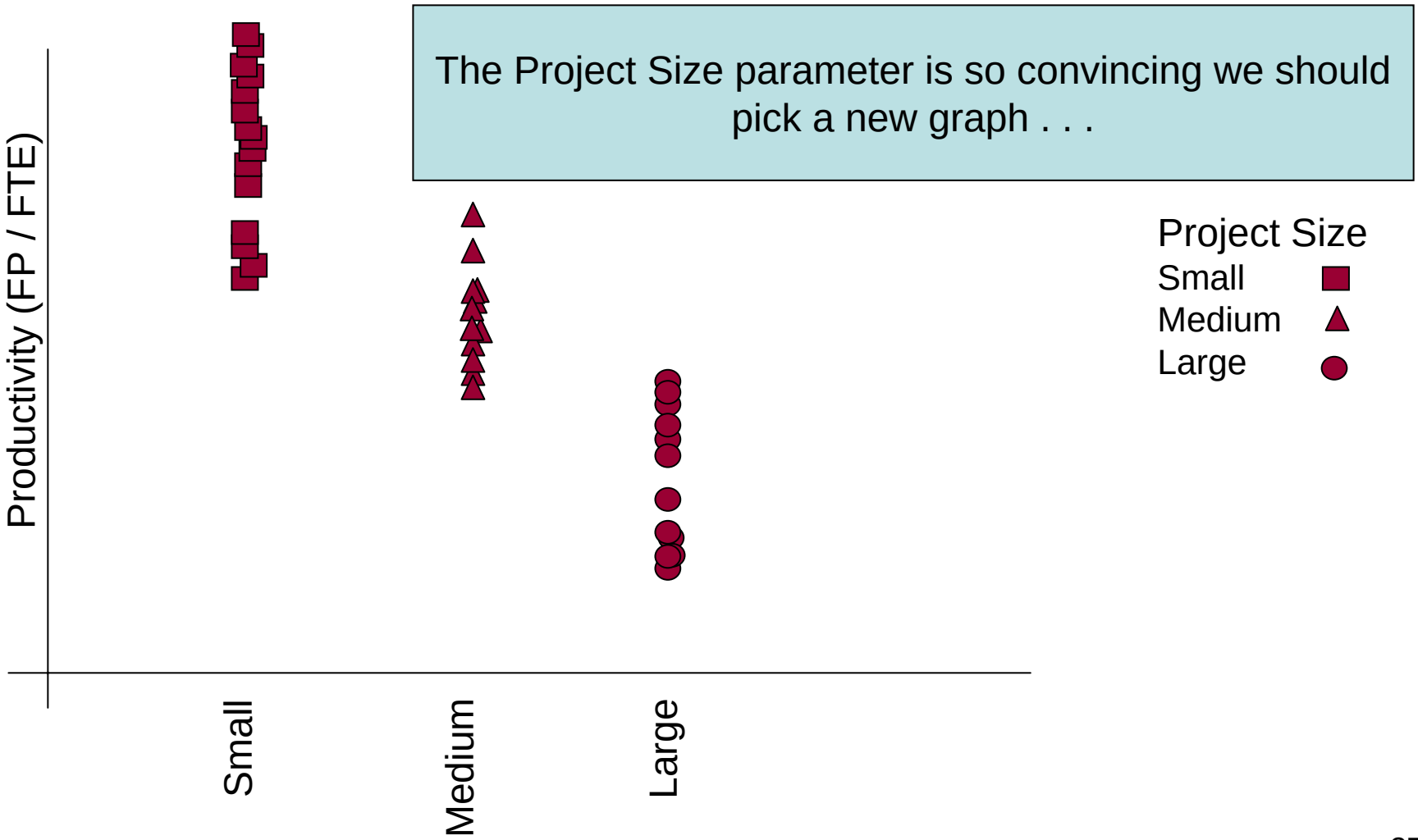
Understanding Variation, my way . . .



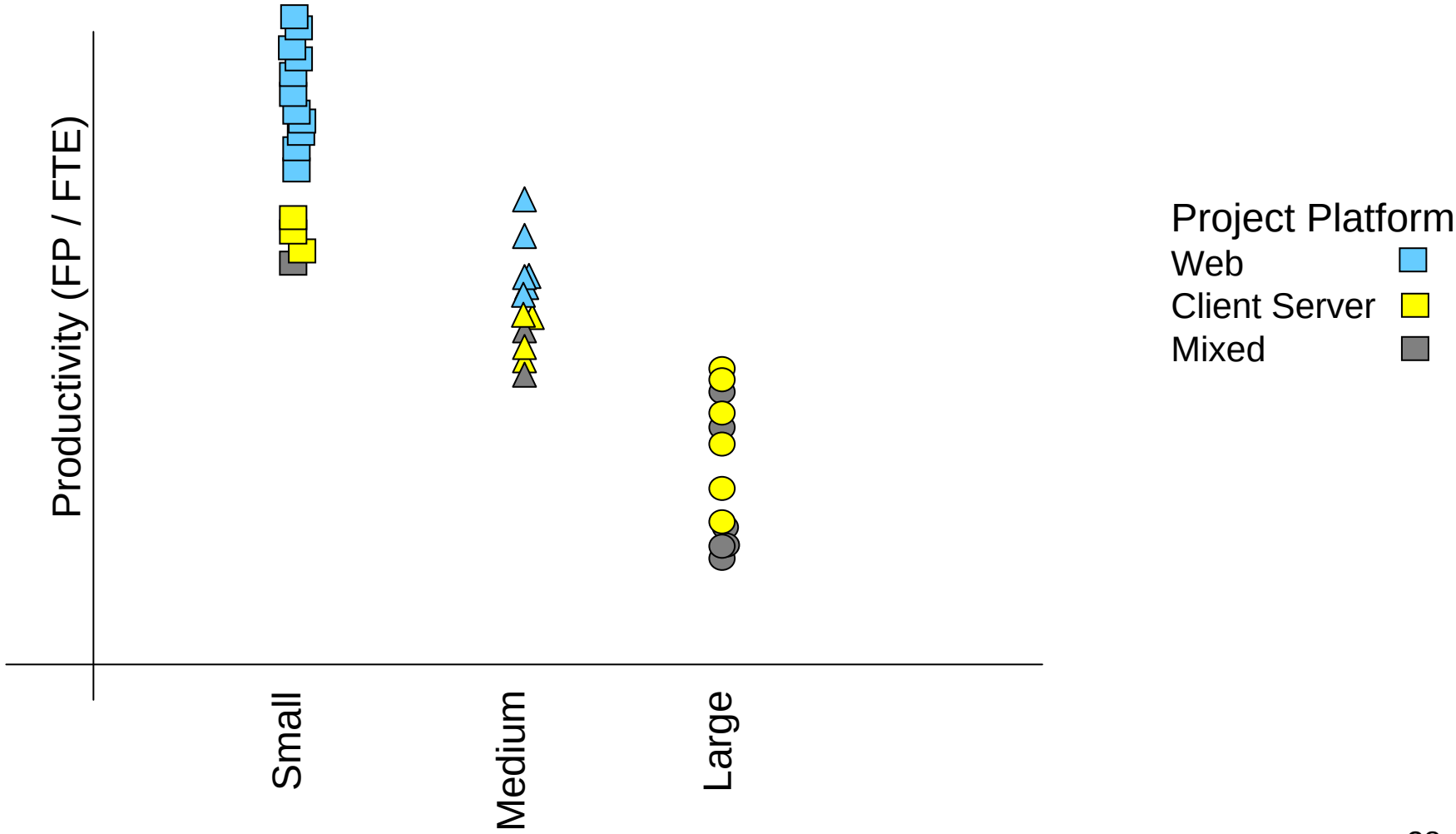
Understanding Variation, my way . . .



Understanding Variation, my way . . .

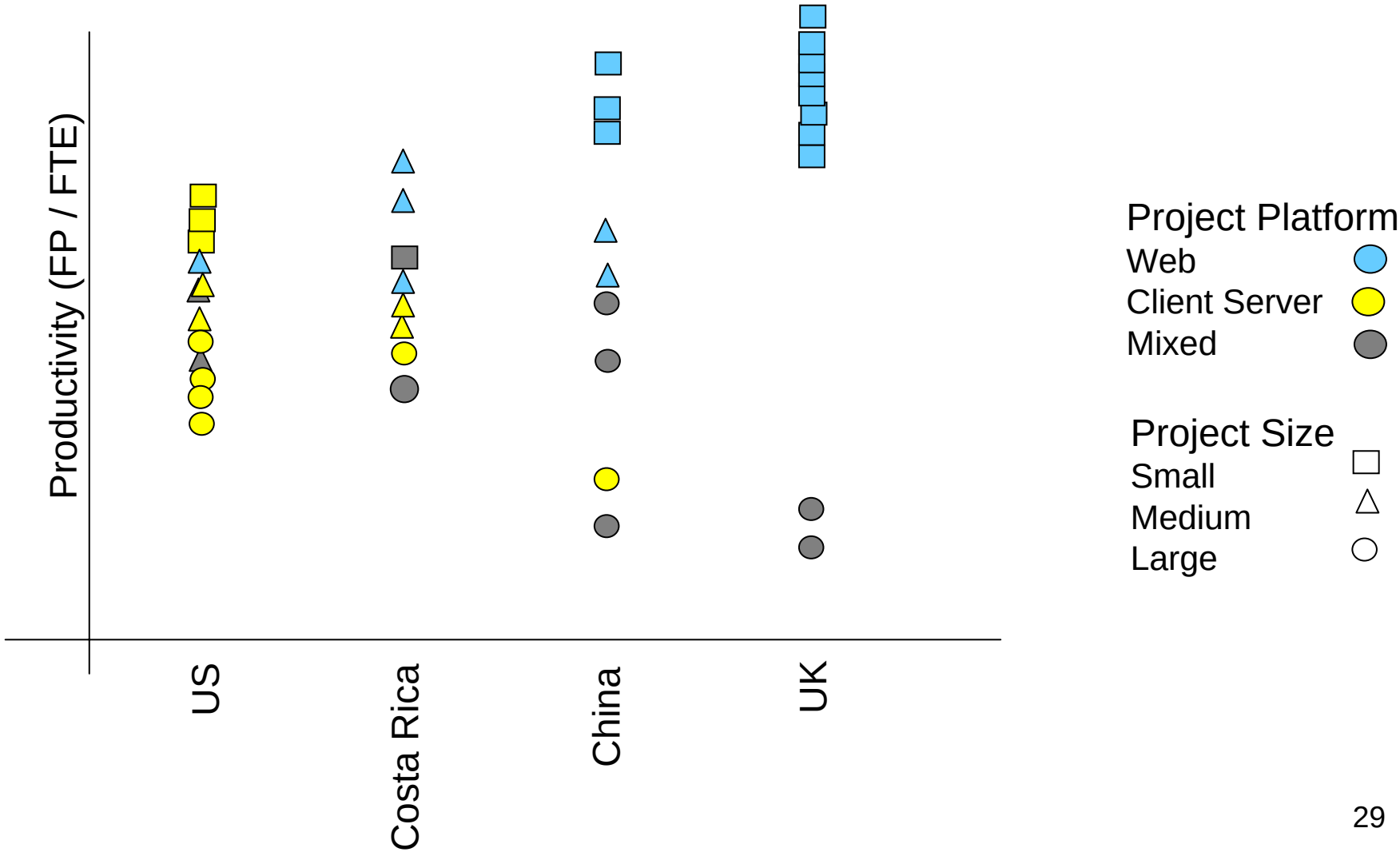


Understanding Variation, my way . . .



This puts a new spin on the original claim regarding UK productivity.

Understanding Variation, my way . . .

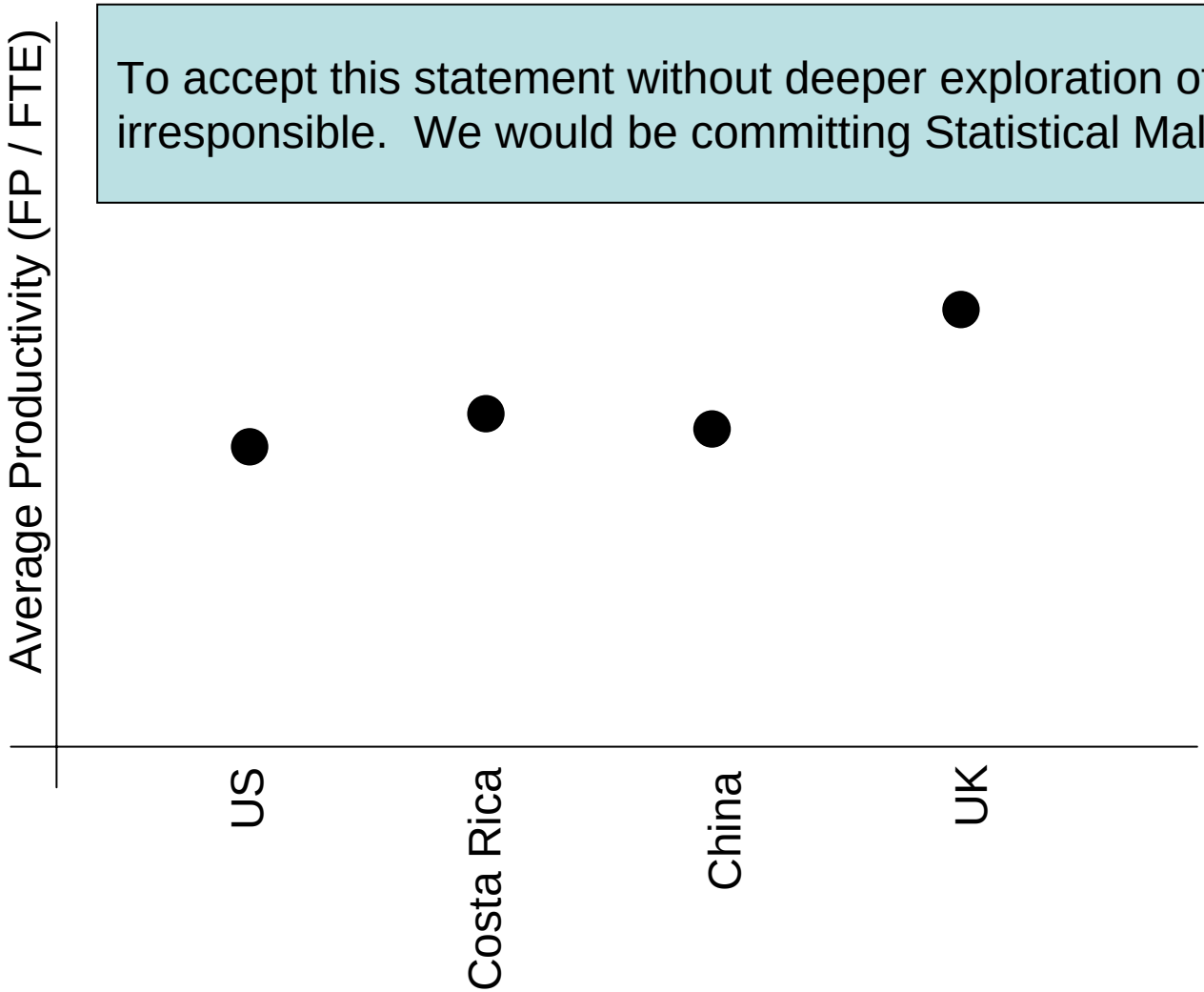


We know we are productive with small web-based projects.

The fact that most of the projects in the UK data are small and web-based more readily explains the apparent higher productivity seen in this far right hand data.

Our project Productivity is highest in UK.

To accept this statement without deeper exploration of the data is irresponsible. We would be committing Statistical Malpractice!!



Statistical Malpractice?

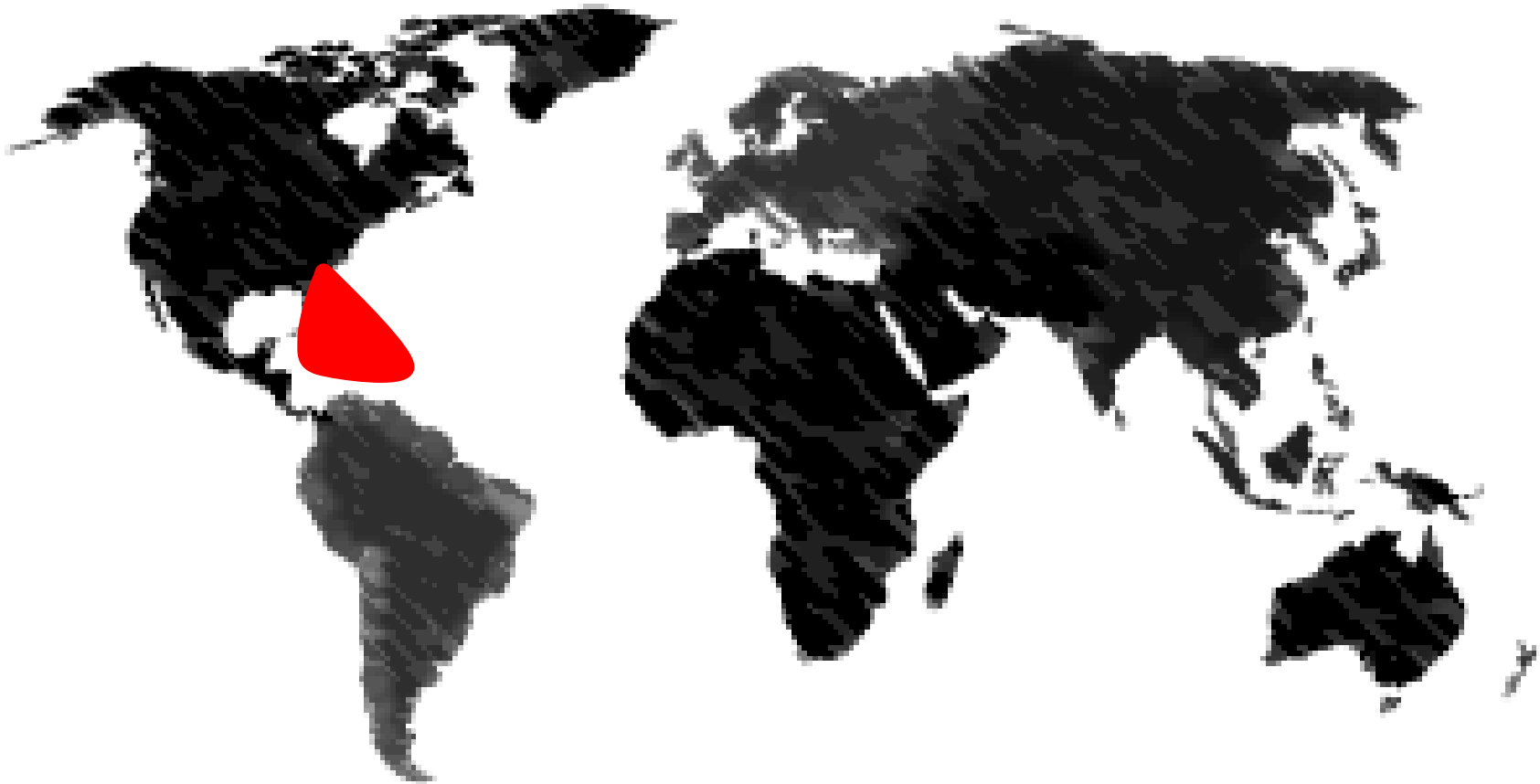
- There is one area of the world's oceans that has claimed more ships than any other.
- Many of these ships have disappeared without a single trace.
- This area, defined by Ft. Lauderdale, San Juan and Bermuda is known as the Bermuda Triangle.
- The Bermuda Triangle is a natural phenomenon that swallows up ships and passengers without a trace. ← The message; are you convinced?

What message is the presentation trying to convey?

- There is one area of the world's oceans that has claimed more ships than any other. ←TRUE
- Many of these ships have disappeared without a single trace. ←TRUE
- This area, define by Ft. Lauderdale, San Juan and Bermuda is known as the Bermuda Triangle. ←TRUE
- The Bermuda Triangle is a natural phenomenon that swallows up ships and passengers without a trace.

- This message is reinforced by the following graph.

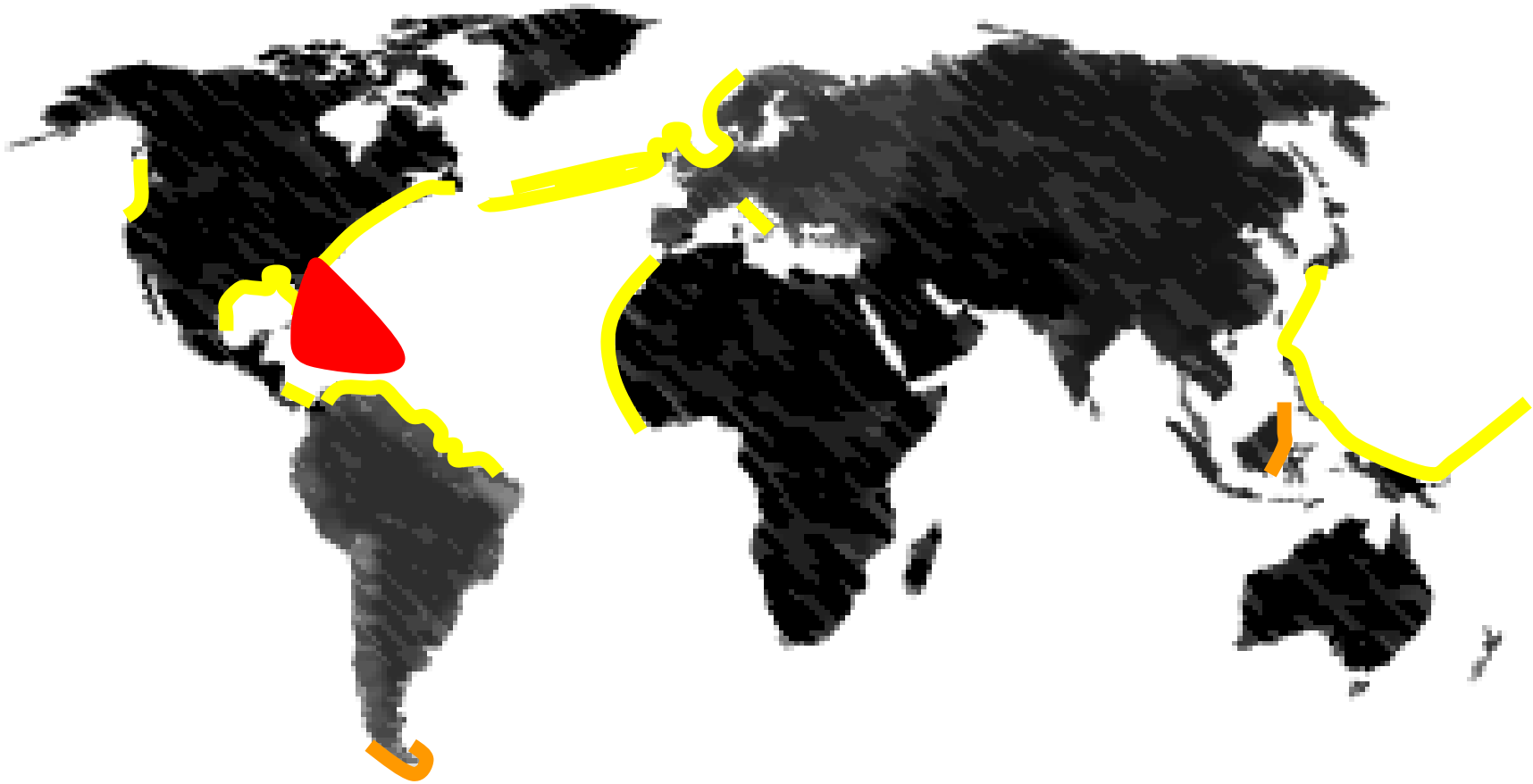
Shipping Losses in the World's Oceans Since 1655



Legend: 5+ vessels lost per year since 1655

Remember: We have to understand the variation!!

Shipping Losses in the World's Oceans Since 1655



Legend:
- 5+ vessels lost per year since 1655
- 2 – 4 vessels lost per year since 1655
- 0.1 – 1 vessels lost per year since 1655
- <0.1 vessels lost per year since 1655

Some oft-omitted facts . . .

- 1/3 of all registered ocean-going vessels on earth are/were registered within this triangle.
- 2/3 of all ocean-going vessels that ever sailed crossed within this triangle at one time or another.

This is a busy part of the ocean

- There are 3500 islands within this triangle, most with barrier reefs.
- The coral in this triangle is the fastest growing on earth.
- The triangle is in Hurricane Alley.

And dangerous too!!

Does the analysis *support* this message?

What we learned from our analysis

- Some basic statistical analysis – we looked at the variation and sought explanation - casts serious doubt on the conspiracy on which 17 “non-fiction” books, 3 movies and countless documentaries have been based.

- Given the volume of shipping traffic and the inherently dangerous islands, reefs and weather in the area, we would be surprised if the Bermuda Triangle did not have a very high rate of shipping loss.

- When a ship and all souls are lost at sea, there are no witnesses. It disappeared.

Blocking

Blocking (Rational Sub-grouping)

- Blocking correctly, incorrectly or not at all can either cause or prevent statistical malpractice.
  - More correctly, this can be the difference between a meaningful analysis and a meaningless analysis.
- Blocking – rational sub-grouping – partitions your data into homogeneous subpopulations.
- Blocking is most commonly recognized as making “apples to apples” comparisons.

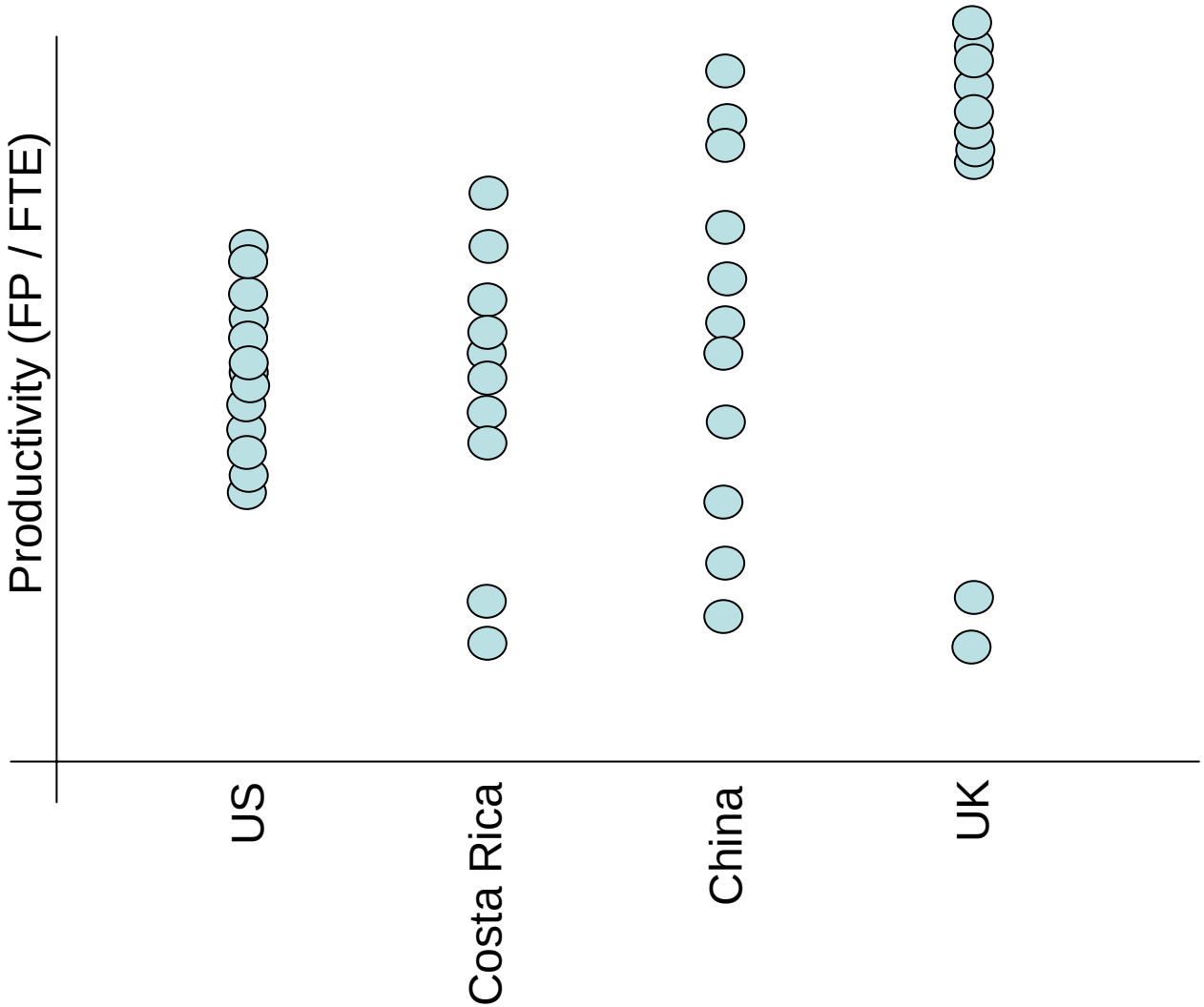
Blocking (Rational Sub-grouping)

- Blocking is the statistical tool that is “make or break” for an analysis.

e.g. Estimating the heights of people in this room and using the information to predict the height of the next person to walk in. This exercise would make more sense to do by gender? By ethnicity?

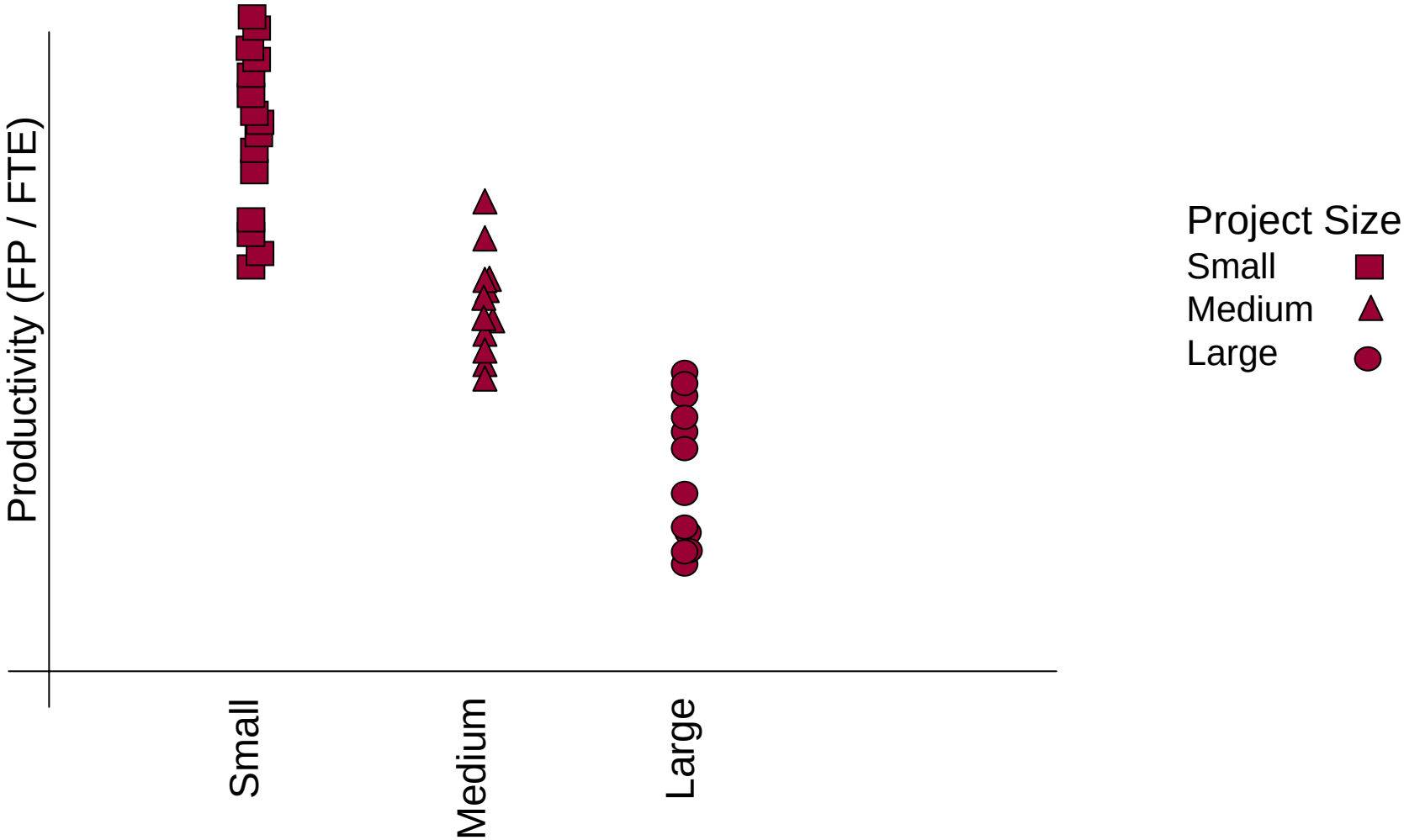
Is country a valid blocking factor?

Critical to block? Bad idea to block?



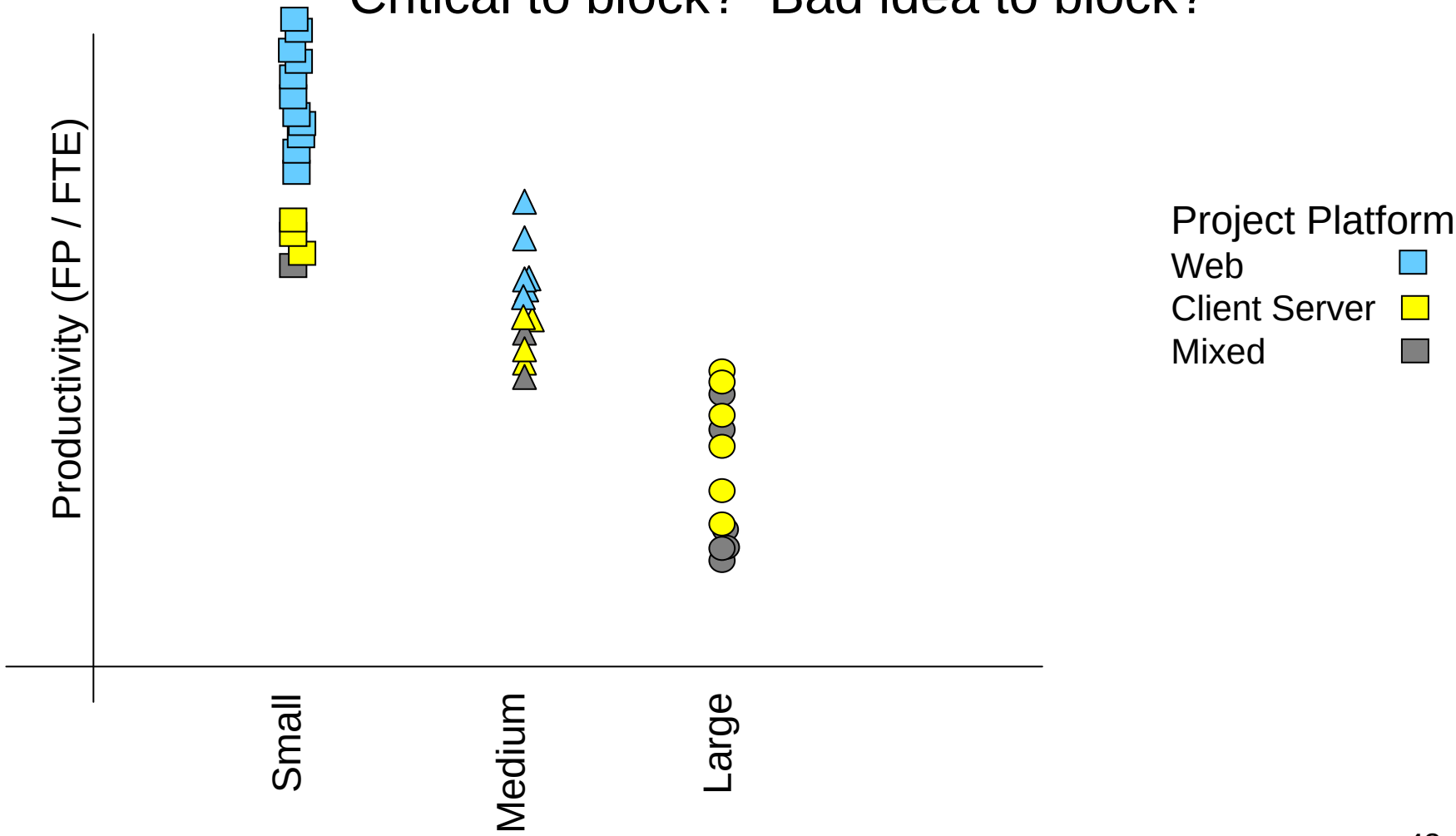
Is project size a valid blocking factor?

Critical to block? Bad idea to block?



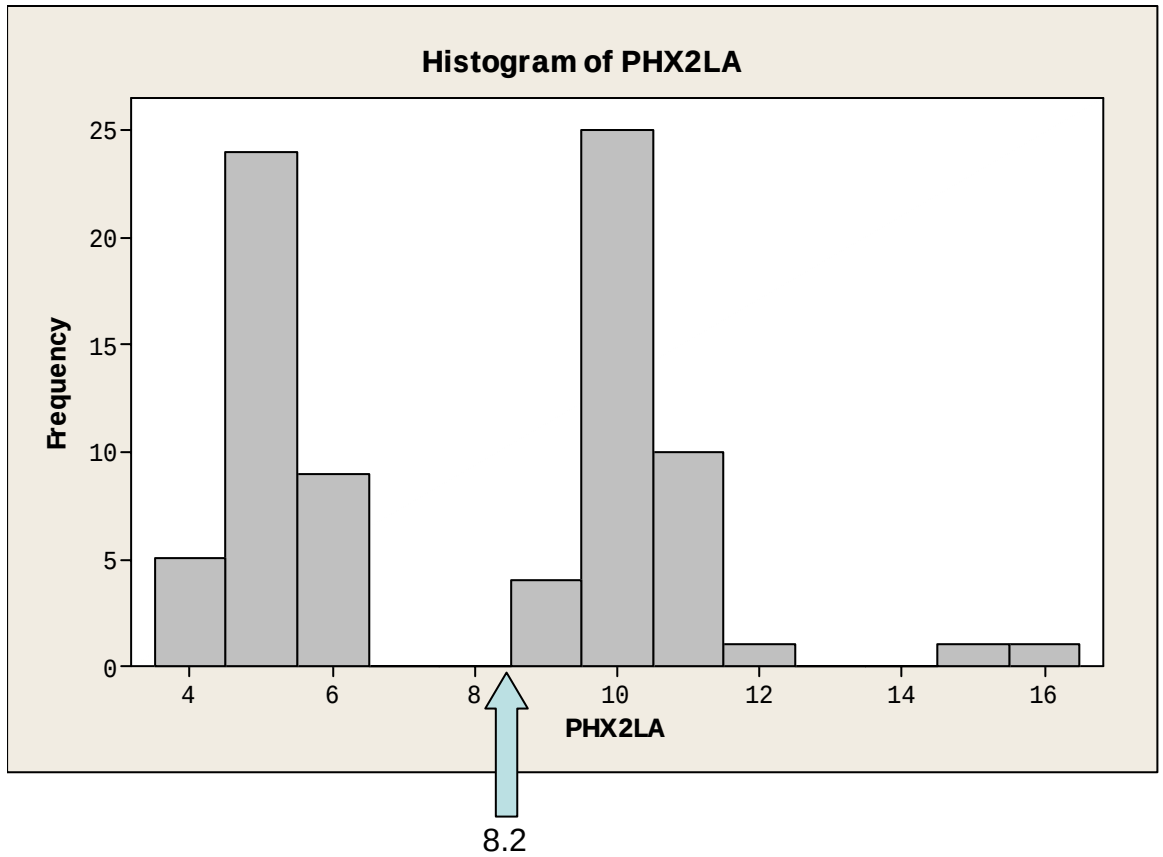
Is project size and platform a valid blocking factor?

Critical to block? Bad idea to block?



Another example . . .

Consider a trucking company that is making hauls from Phoenix to Los Angeles. Can we use MPG data to forecast fuel expenses for the next quarter?



Average MPG = 8.2

It would be irresponsible to make inferences from this data without understanding the nature of the data we have at hand.

Note that we have to assume fuel costs per gallon remains constant.

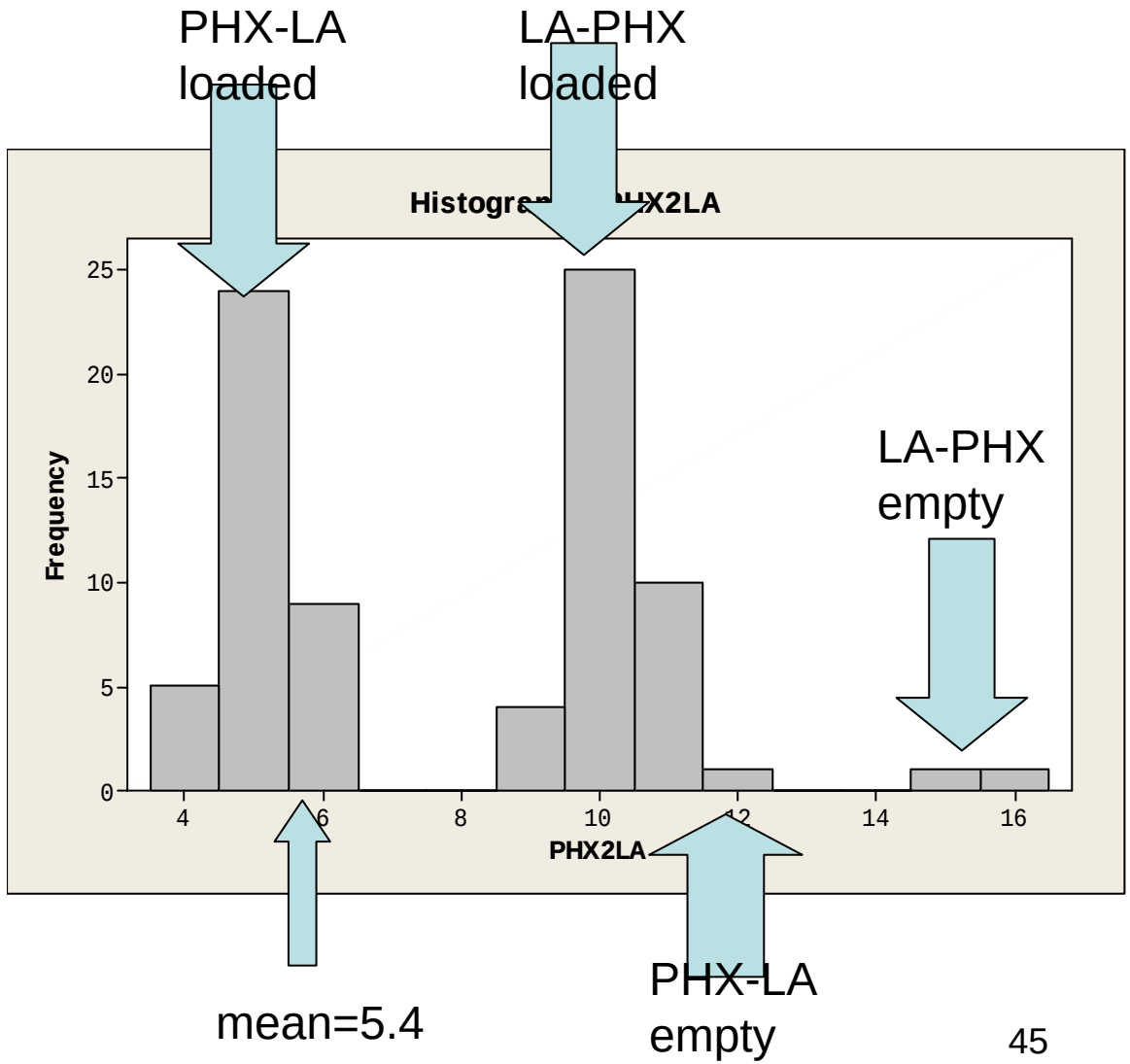
A more interesting example . . .

Because we understand this data, we know we have 4 data sets.

The most intelligent inferences can be made by understanding your data to this level, at least.

For example, predicting fuel expenses can be done by estimating the number of each type of trip and its average fuel consumption and combining.

The number of each trip type is a higher resolution of data than just number of miles.



Three approaches to Blocking

- Try it and see if it looks right. ← demonstrated earlier.
- Put some thought into it (e.g., productivity for web projects should be "close cousins").*
- There are some basic analyses to ascertain whether we blocked correctly or unnecessarily. There are no red flags in an analysis that says where blocking should be done

*We have so much confidence in this "should" that we would be forced to investigate if this were not the case.

Statistical Inference

From Summary to Inference

- The science of statistics gives us the best known methods for summarizing data.
- Ultimately the value found in statistics is through inferences.
- We are most familiar with statistical inferences through project forecasting.
- There are places where we are even more familiar with statistical inferences.

We are all familiar with confidence intervals

- 42% support candidate D
- 49% support candidate R
- margin of error +/-3%

Some definitions . . .

(Easier to give an example)

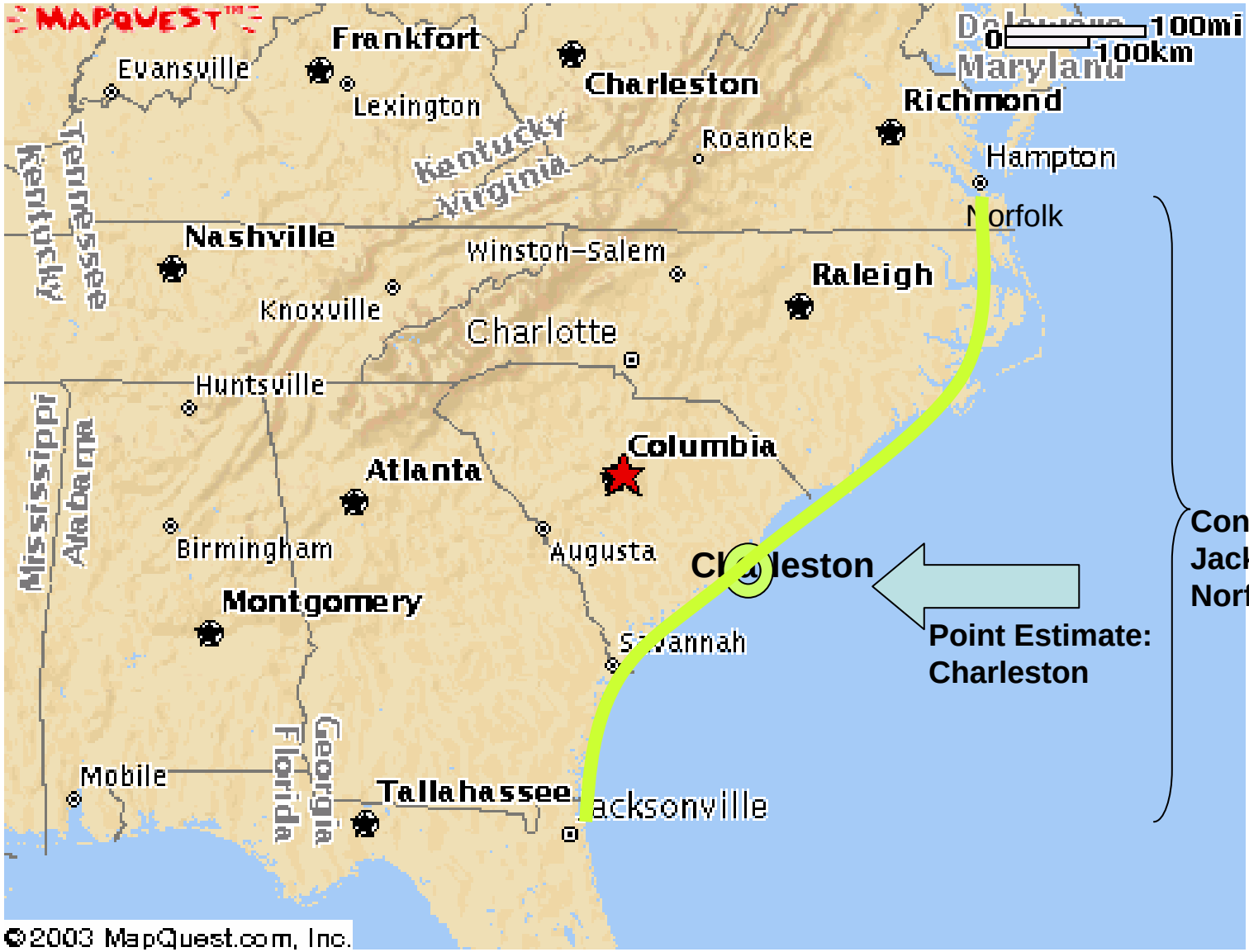
- Point estimate
  - My team produced an average of 30 function points per person month for the last 14 small web projects. So I expect us to produce about 30 function points per person month on the next project.
  - Note: *point estimate* is usually an average. *Average* is also commonly called an *expected value* or *expectation*.
  - Another point estimate you are used to: Hurricane Isabel is expected to make landfall at Charleston.

Some definitions . . .

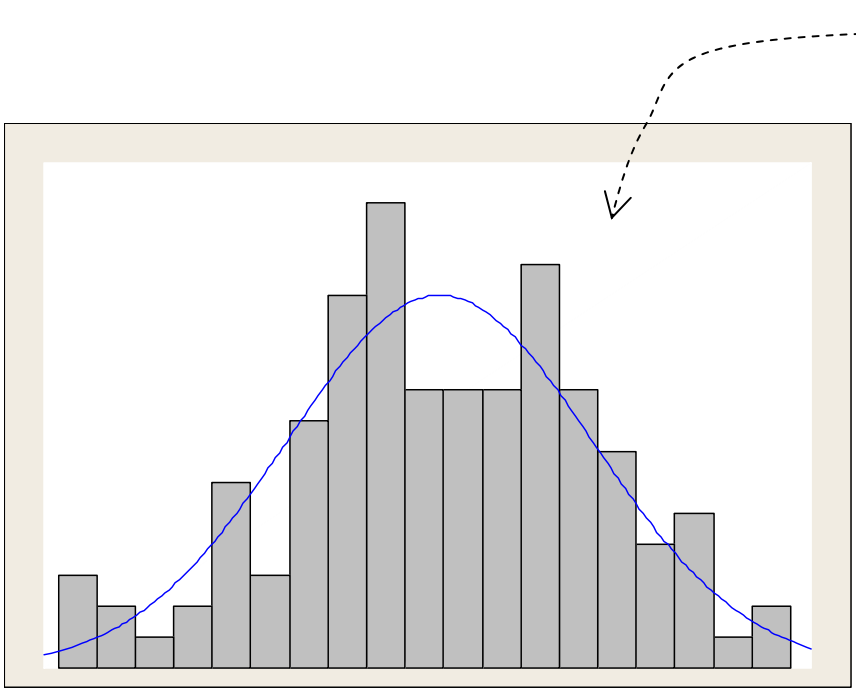
(Easier to give an example)

- Confidence Interval
  - I expect us to produce 30 fp/pm on our next project. I am 95% confident that we will produce between 10 and 78.
  - This is the foundation for what is known as statistical inference. We are beyond analyzing data. We are using it for "predictive" purposes.
  - *Any time we have enough data to calculate a point estimate, we can calculate a confidence interval; which can be enormously more useful.*
  - Another confidence interval you are used to: Hurricane Isabel will make landfall between Norfolk and Jacksonville, with Charleston as most likely to experience the eye.

Two ways to predict landfall for Hurricane Isabel: a point estimate or a confidence interval. Which is more likely to represent landfall correctly?



The essence of the confidence interval is--based on previous data we have sampled--predicting where additional data is likely to fall.



95% Confidence Interval

The 95% confidence interval for hurricane Isabel's landfall is between Jacksonville and Norfolk.

Exactly Speaking: Of the last 100 hurricanes that matched Isabel's characteristics, 95 of them made landfall between these two cities.

Common Language: There is a 95% chance that hurricane Isabel will make landfall between these two cities.

Confidence Intervals and Sample Size

- If we are only able to consider a few past hurricanes, our 95% confidence interval will be pretty wide. If we can consider many, many hurricanes, our 95% confidence interval will be small.

- The larger a sample size, the smaller the confidence interval.

Geek Zone

- If you want to take it to extremes, for a sample size of 1, the confidence interval is +/- infinity. With an infinite sample size, the confidence interval is +/- 0.
  - If you only have one mid-Atlantic hurricane to draw experiences from, you can be confident it will make landfall somewhere between Newfoundland and Tierra del Fuego.
  - If you knew everything about mid-Atlantic hurricanes, you could know exactly which beach it will land on.

Let's link confidence intervals and sample sizes . . .

- My team produced an average of 30 function points per person month for the last 14 small web projects. So I expect us to produce about 30 function points per person month on the next project.
- I expect us to produce 30 fp/pm on our next project. I am 95% confident that we will produce between 10 and 78
- Suppose I randomly selected only 7 of my team's projects and computed a 95% confidence interval. It would be from 0 fp/mm to 108 fp/mm.
- Suppose I had 20 small web projects worth of experience and computed a 95% confidence interval. It would be from 18fp/mm to 62 fp/mm.
- Notice how my confidence interval gets smaller with a larger sample size, and vice versa.

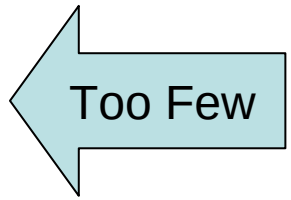
How big does my sample size need to be?

How small do you need your confidence interval to be?

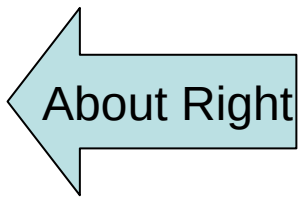
Reminders

- I expect us to produce 30 fp/pm on our next project. I am 95% confident that we will produce between 10 and 78.

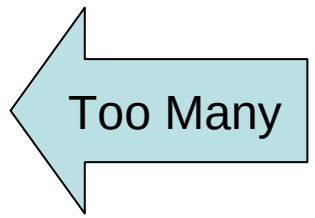
- It is easy to imagine a confidence interval that is too big to be useful. With a sample size of 2 projects, my 95% confidence interval would be 0fp/mm to 1000fp/mm. duh.



- Based on my 14 projects, I have computed a 95% confidence interval from 10 fp/mm to 78 fp/mm.



- A sample size that is too big might give you a 95% confidence interval like 32.14fp/mm to 32.17fp/mm.



How big does my sample size need to be?

How small do you need your confidence interval to be?

A general sample size formula is: $n = (2s/d)^2$

Where *d* is the width of the confidence interval we are able to accept. And *s* is the standard deviation of our data.

A contradiction has appeared. We need a sample to compute a standard deviation and we need a standard deviation to identify a sample size.

How small does my confidence interval need to be?

- How many takeoff-landing cycles can a 747 airframe take before cracks start forming in the airframe?
  - 12000 +/- 250
  - 12000 +/- 2500
- How many houses does a girl scout have to visit to sell 20 boxes of girl scout cookies?
  - 45 +/- 2
  - 45 +/- 10

Computing Sample Sizes

http://www.stat.uiowa.edu/~rlenth/Power/index.html ←there are several options

X

Edit View Favorites Tools Help

Back Search Favorites

http://www.stat.uiowa.edu/~rlenth/Power/index.html

Java applets for power and sample size

Select the analysis to be used in your study:

- CI for one proportion
- Test of one proportion
- Test comparing two proportions
- CI for one mean
- One-sample t test (or paired t)
- Two-sample t test (pooled or Satterthwaite)
- Linear regression
- Balanced ANOVA (any model)
- Generic chi-square test
- Generic Poisson test
- Pilot study

Run selection

This software is intended to be useful in planning statistical studies. It is not intended to be used for analysis of data that have already been collected.

Each selection provides a graphical interface for studying the power of one or more tests. They include sliders (convertible to number-entry fields) for varying parameters, and a simple provision for graphing one variable against another.

Each dialog window also offers a Help menu. Please read the Help menus before contacting me with questions.

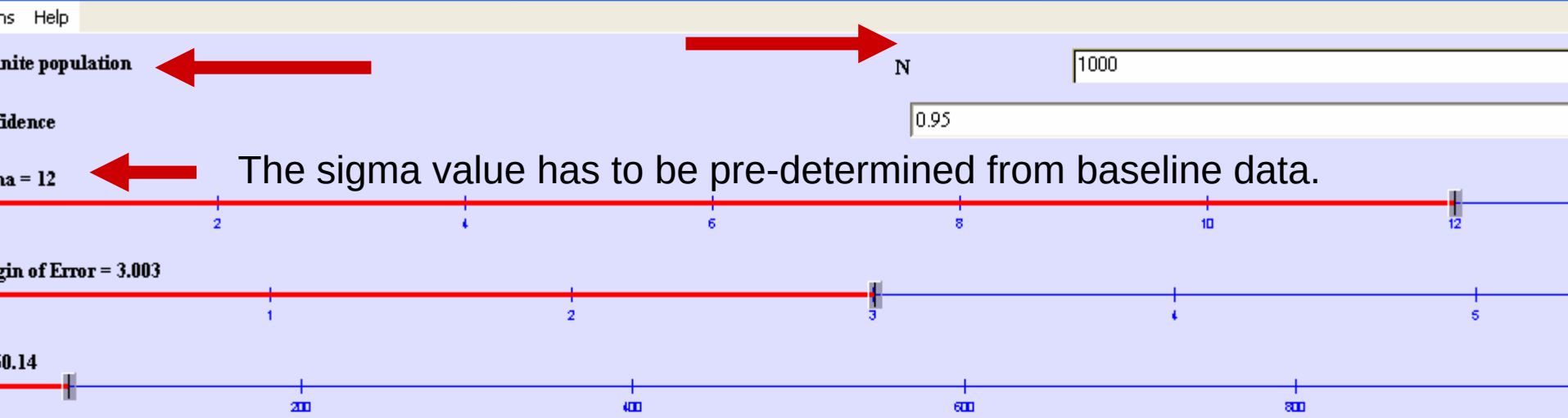
The "Balanced ANOVA" selection provides another dialog with a list of several popular experimental designs, plus a provision for specifying your own model.

Note: The dialogs open in separate windows. If you're running this on an Apple Macintosh, the applets' menus are added to the *screen* menubar -- so, for example, you'll have two "Help" menus there!

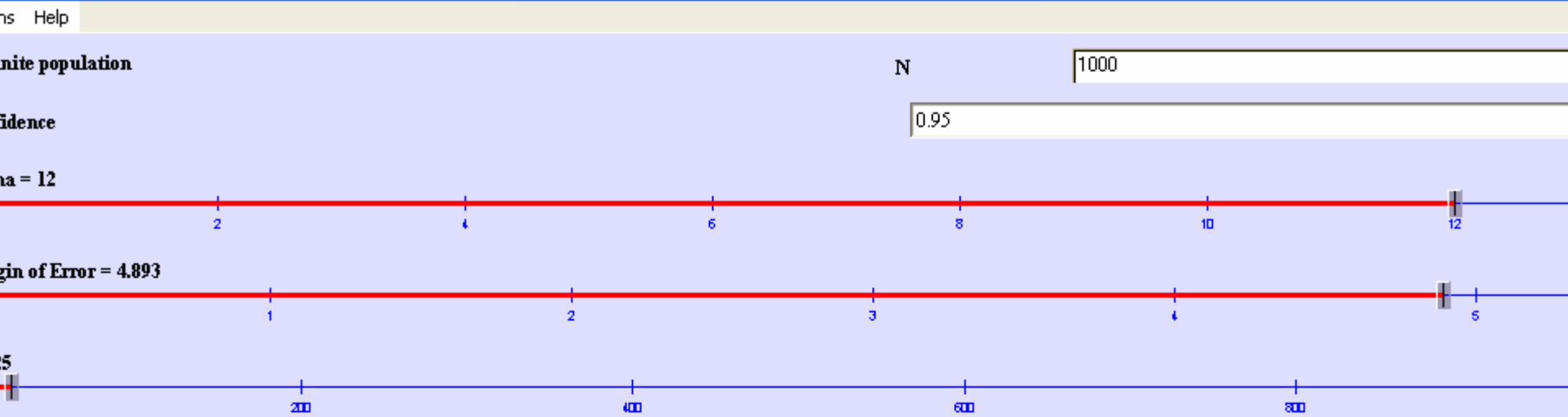
You may also download this software to run it on your own PC.

te: These require a web browser capable of running Java applets (version 1.1 or higher). If you do not see a selection list above, chances are that you either have disabled Java, ur browser is not new enough., or you need to download a JRE plug-in from java.sun.com. Due to a compatibility bug, many plug-ins size the applet window before allowing for an ditional strip with a security warning.; drag the bottom of the window downward a bit to compensate.

Internet



If we need a margin of error of 3, we need a sample size of 60.
If our baseline data does not have 60 pieces, we cannot achieve a +/- 3 margin of error.



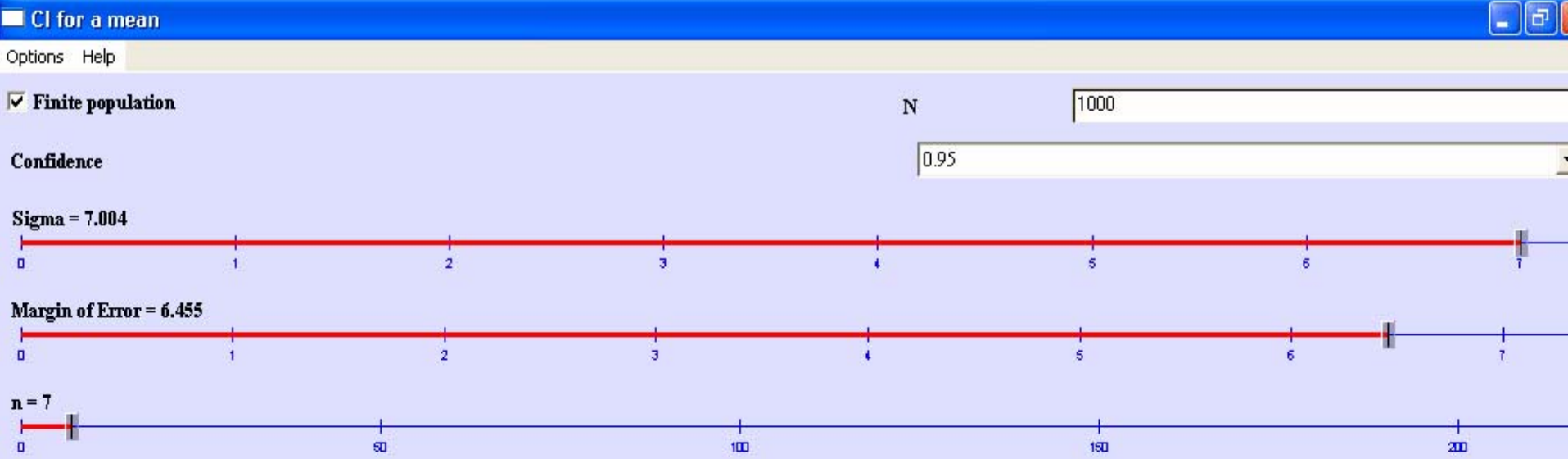
If we only had 25 projects in our baseline, we could only expect a +/- 4.9 margin of error.

Putting this together…

A team has completed 8 small web projects together in the last year. They have averaged 42 fp/mm with a standard deviation 7 fp/mm.

1) What is the precision with which we can predict the productivity of the next project for this team?

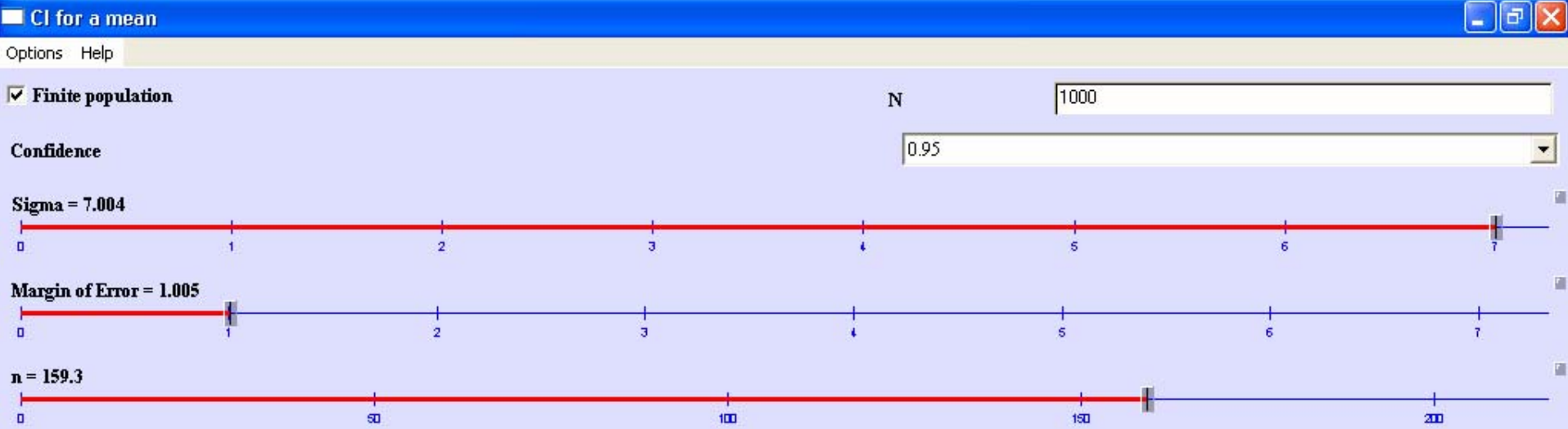
2) How many projects will this team need to complete to start forecasting their productivity to within +/- 1.0 fp/mm?



A team has completed 8 small web projects together in the last year. They have averaged 42 fp/mm with a standard deviation 7 fp/mm.

1) What is the precision with which we can predict the productivity of the next project for this team?

We are 95% confident the next project completed by this team will show productivity of 42 +/- 6.5 fp/mm
= between 35.5 fp/mm to 48.5 fp/mm



A team has completed 8 small web projects together in the last year. They have averaged 42 fp/mm with a standard deviation 7 fp/mm.

2) How many projects will this team need to complete to start forecasting their productivity to within +/- 1.0 fp/mm?

We would need a baseline of 160 similar projects to be able to forecast to within +/- 1 fp/mm.

In Summary . . .

- We all have a problem solving toolbox.
- Enriching your toolbox with Statistical Thinking is really valuable.
- Your insisting on seeing statistically valid presentations is very healthy.
- You should neither commit statistical malpractice nor become a statistical victim.

Definition: *Statistical Malpractice*

- The act of misleading, misinforming or confusing others by means of an incomplete or erroneous statistical analysis of data, whether due to intent or accident.

Definition: *Statistical Victim*

- One who as been misinformed, misled or confused by an incomplete or erroneous statistical analysis. Also know as *Victim.*