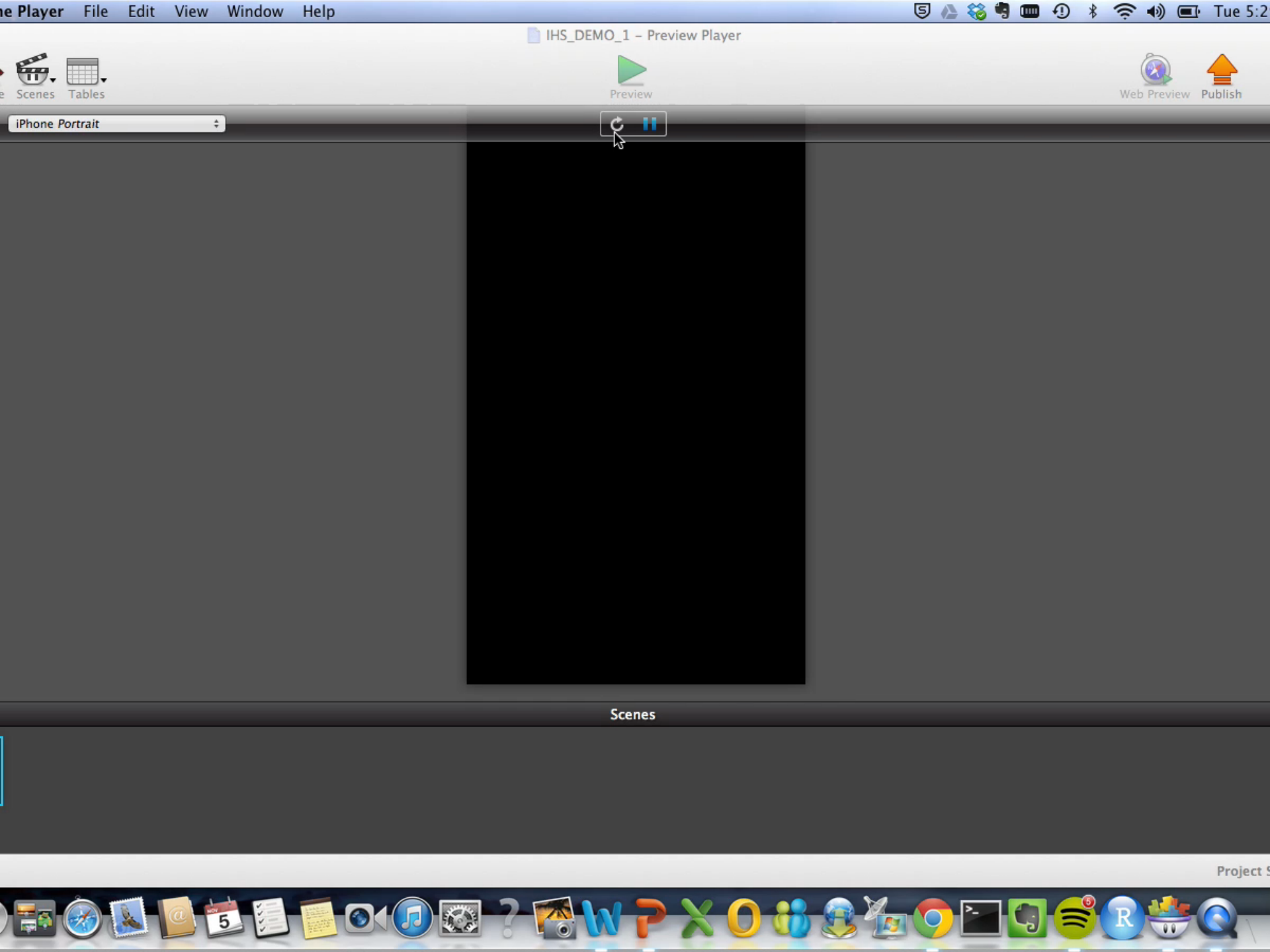
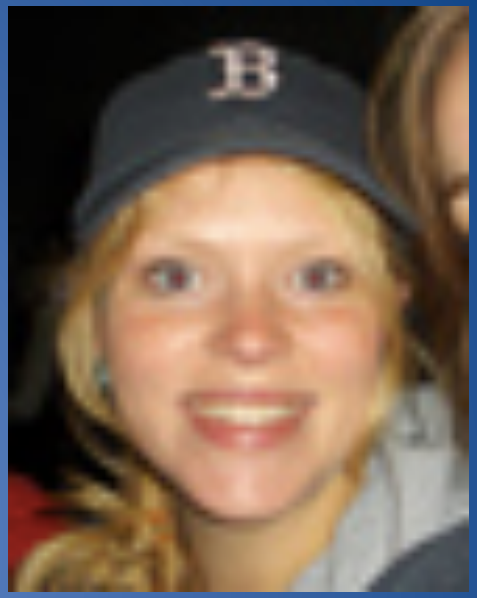
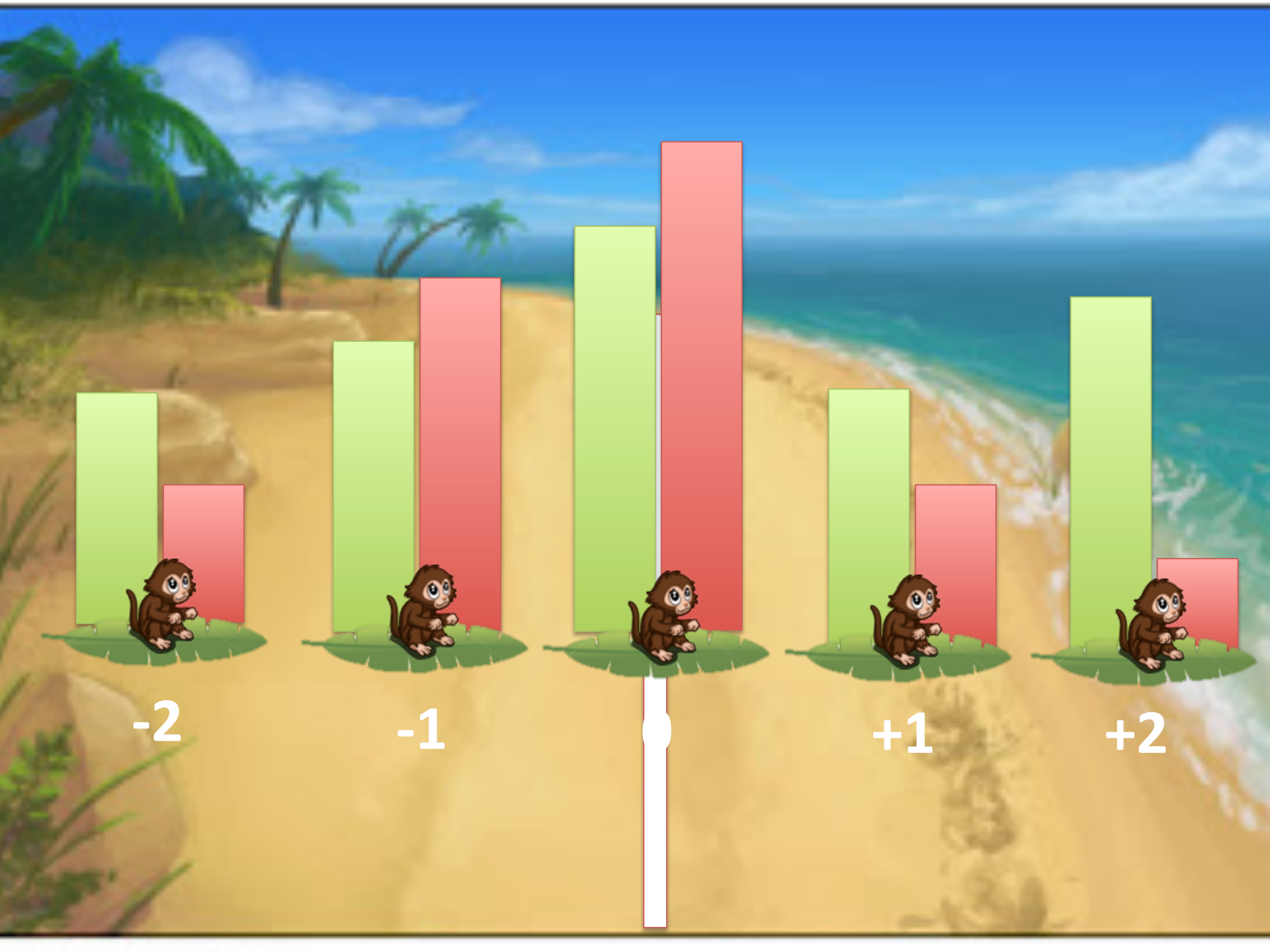
I HATE
STATISTICS

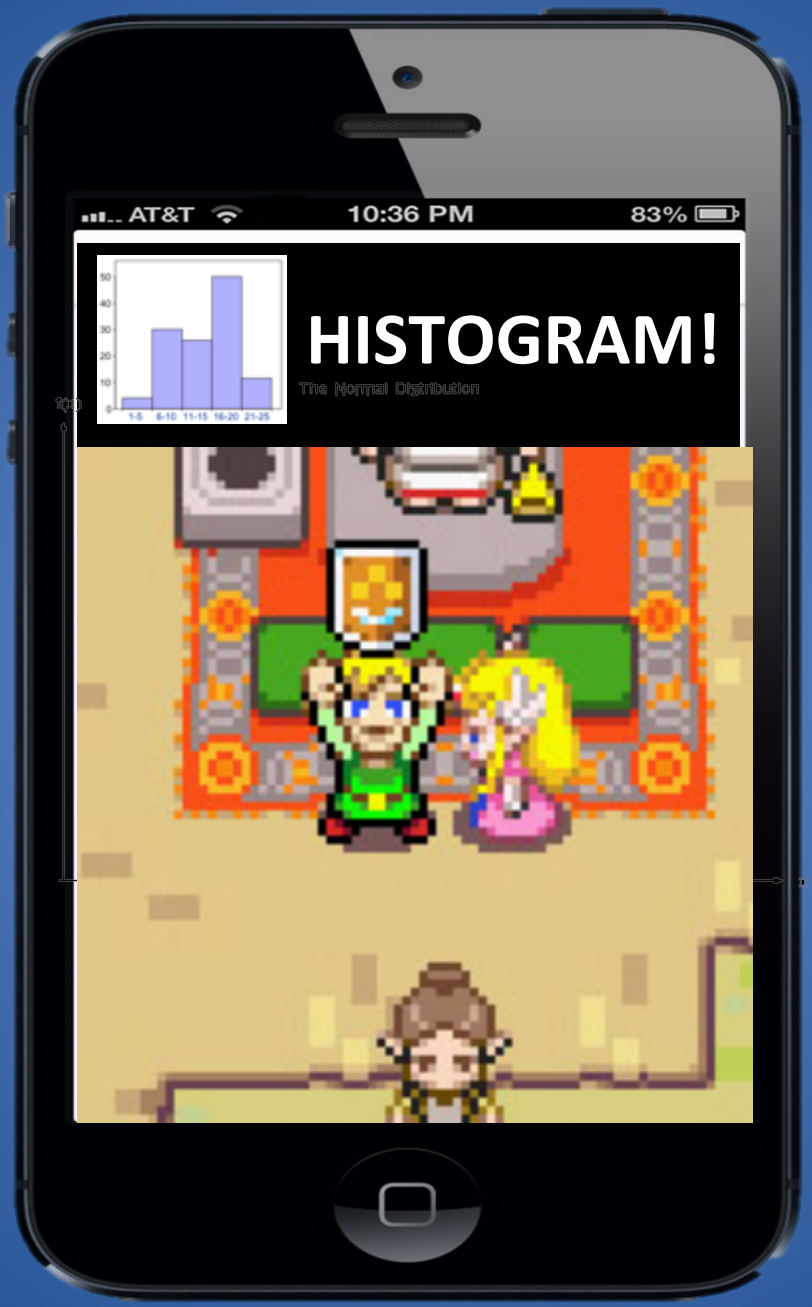

Player File Edit View Window Help
Tue 5:2
IHS_DEMO_1 – Preview Player
Scenes
Tables
Preview
Web Preview
Publish
iPhone Portrait
Scenes
Project S

VS



WHO WON?


-2
-1
+1
+2


AT&T
10:36 PM
83%
HISTOGRAM!
The Normal Distribution

Section 2

REGRESSION

Introduction

So far we have concerned ourselves with sets of data in which only one reading on each of a given set of objects has been recorded e.g. number of peas/pod, I.Q.'s, children's weights, etc.

In this section we will consider sets of data in which there are readings on more than one variable. The following are examples of such sets of data.

I. In an experiment of the capacity of electrolytic cells, four cells were taken and filled with 0.50, 0.55, 0.60 and 0.65 ml of electrolyte respectively. Each cell was charged up at constant current to estimate the capacity and this was repeated ten times on each cell.

The results were as follows:-

Quantity of electrolyte (ml)	0.50	0.55	0.60	0.65
Mean capacity	131.27	133.00	134.55	137.01

II.

Number of pupils per teacher	17	19	12	10	16	13	18	20	22	13	19	8
Percentage of pupils passing 3 GCE subjects	5	6	8	7	5	7	3	4	5	9	3	11

We can plot such sets of data on a graph called a SCATTER DIAGRAM. The scatter diagrams for the two examples given are shown on p. 20.

It can immediately be seen that in each example the points are closely clustered around a straight line. This indicates that a relationship may exist between the variables in each example i.e. between quantity of electrolyte and mean capacity, and between the number of pupils/teacher and the percentage of pupils passing three GCE subjects. The line around which the points cluster is called a REGRESSION LINE. Suppose we have two variables x and y; using a regression line we can, given any value of the variable x,

J. Hine et al., *A Programmed Test in Statistics Book 4: Tests on Variance and Regression*
© J. Hine and G. B. Wetherill 1975

VS