

## Checkpoint 2:

These two figures are a rough mock up of our final dash board. There are several components involved. The main figure at the top of the dash board is the visualization of the “workout groups”. After we run our clustering algorithm on our data, we reduce the vectors to a 2 dimensional space (as opposed to 5) and plot the output. We’ve added a few features and still have more in production as to how you interact with the data.

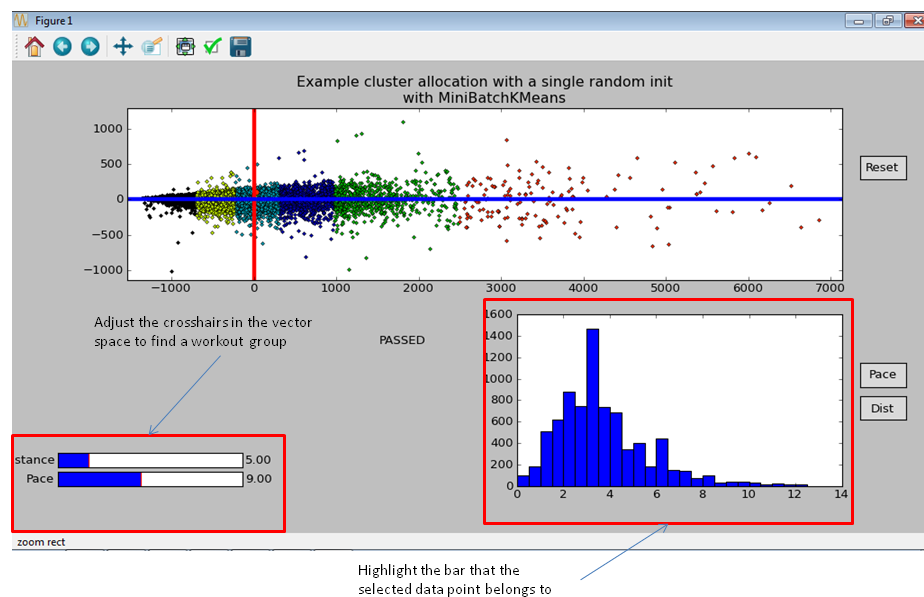


Figure 1 – Depicts a running profile adjuster and running statistics

## Interaction:

Our goal is to understand the properties of every clustering our visualization. Since every point represents an athlete, we have implemented a way of selecting a data point in the plot. With this capability we would like to highlight what bar the athlete fits in for various statistics, view their pace pattern (if available), or the nominal data associated with their workout (emotion, terrain, weather, etc.). We also want to allow the user to generate a running profile to see what groups they would best fit in using the knobs in the bottom left corner. The bottom left of figure 1 shows a few knobs that can be adjusted to generate workout a vector. With this vector we can then plot it to see which group it fits best with. The crosshairs in figure 1 and 2 will show where your choices for the knobs cause you to be in the clustering

## Special Topics in: Data Science

### Fit-R-Us

Jason Bolden and David Spellman

The user can also zoom in or out of the workout groups figure. To return back to the original zoom level they simply need to press the reset button. Since matplotlib begins to slow down when more data points are present, we will only plot the cluster centers at the original level. The user can then click on a center and the figure will resize to fit and plot only the athletes in that group. We would also like to display statistics of a group as a whole when selected (i.e. highlight the range of bars that athletes occupy in a histogram).

We would also like to implement options to cluster on varying features. We intend to add options to specify which feature to include in the vector space (average pace, duration of run, distance, calories burned, etc.). We would also like to add the option to cluster based on pace pattern. Most of the runs have a vector of their current speed sampled every 10 second throughout their run. We will use the temporal correlation combined with a distance metric when clustering by pace pattern.

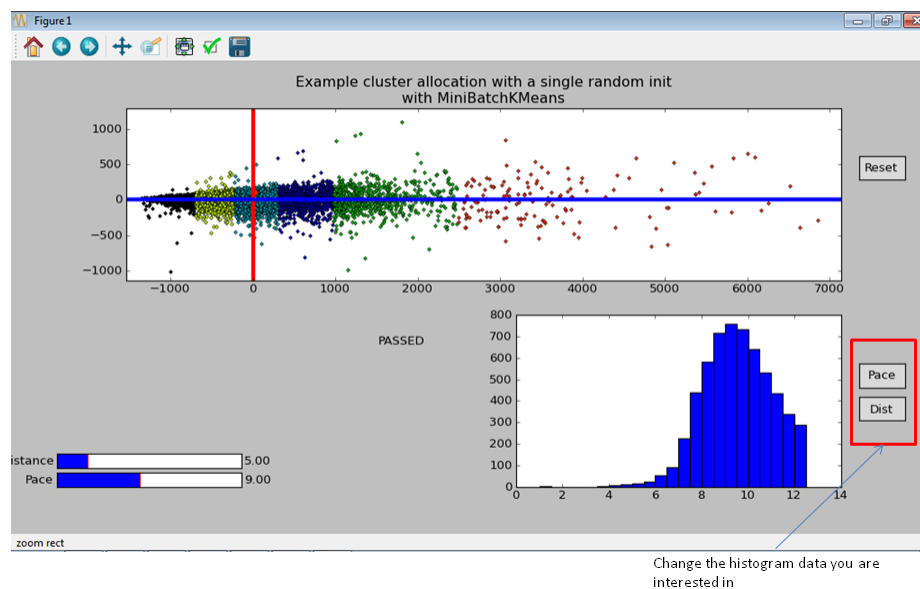


Figure 2 – Options to select which statistics are displayed