

# A SIMPLE VISUALIZATION OF THE STANDARD AND POORS 500 INDEX

R User 2010 Conference Poster Session, NIST-MD, July 21st, 2010

Nadeem Faiz

Chemical Analysis Group, Agilent Technologies Inc., Wilmington, DE

We seek to answer the following question:

Can we present in a simple visual manner a comparison of the recent S&P 500 Index with its historic performance?

Nadeem is a Mass Spectrometer Manufacturing Engineering Manager at Agilent Technologies Inc.. R is used at Agilent in mining and analysis of mass-spec manufacturing and quality data.

The content of this poster is unrelated to Agilent Technologies products and businesses.

## The S & P 500 Index

The Standard and Poors 500 is an equity index that includes 500 leading companies in leading industries of the US economy. It represents 75% of US equities, is large cap focussed, and is often a proxy for the total market [1]

The S&P 500 Index is a capitalization-weighted index. It is calculated as below

$$S\&P\ 500\ Index\ Level = \sum_{i=Company\ in\ Index}^{500\ companies} \frac{Stock\ Price_i \times Number\ of\ Shares\ of\ Stock_i}{Index\ Divisor} \quad [2]$$

The *Index Divisor* scales the 500 company market value (on 7-2-2010 9.3 T\$) to a more manageable number which is the Index Value (currently 1030). The *Index Divisor* is adjusted to smooth the S&P 500 Index to additions or deletions of companies and to other Index maintenance activities [2]

[1] "S & P 500 Indices", <http://www.standardandpoors.com/indices/sp-500>

[2] "S&P Indices: Index Mathematics Index Methodology", May 11, 2010, Index Services, <https://www.sp-indexdata.com>

## Comparing the Index with Historical Performance

Is the Index on any particular day within the range of historical performance? is it lagging historic performance? or is it leading historical performance?

To answer this question we calculate box plot parameters on the amount of change in the index for a given time period and then compare with the change on that day.

### STEP 1

The relative change in the index is calculated for a range of time intervals. If  $i$  is the

interval in days, relative change on day  $j$   $i \Delta_j$  is as  $i \Delta_j = \frac{Index(day\ j) - Index(day\ j - i\ days)}{i}$

### STEP 2

When repeated for each day in the index the following dataset results:

$$S_i = \{i \Delta^{1-3-1950}, i \Delta^{1-4-1950}, \dots, i \Delta^j, \dots, i \Delta^{7-6-2010}\}$$

### STEP 3

Calculate statistics such as mean, median, 10<sup>th</sup>, 25<sup>th</sup>, 75<sup>th</sup>, and 90<sup>th</sup> percentiles on  $S_i$ .

### STEP 4

Repeat Steps 1 to 2 for several time intervals (in days) {20,50,100,200, 400,800,1000,1200,...,2400} resulting in several datasets { $S_{20}, S_{50}, S_{100}, \dots, S_{2200}, S_{2400}$ }. Calculate statistics as in Step 3 on datasets { $S_{20}, S_{50}, S_{100}, \dots, S_{2200}, S_{2400}$ }.

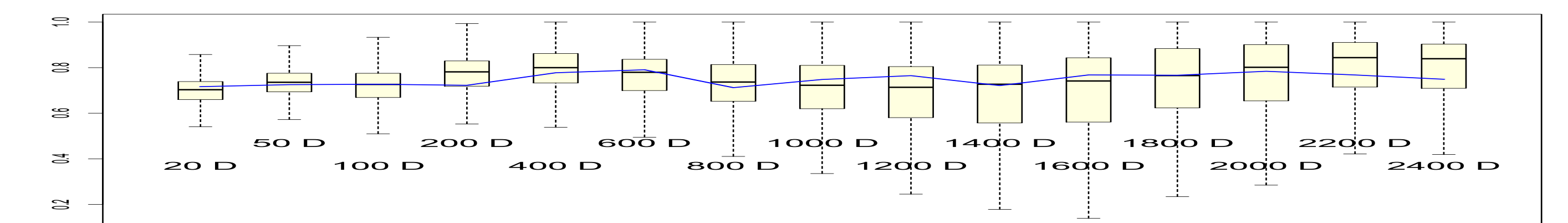
### STEP 5

For the given date calculate the relative change values for all intervals.

$$\{20 \Delta_{7-2-2010}, 50 \Delta_{7-2-2010}, 100 \Delta_{7-2-2010}, \dots, 2200 \Delta_{7-2-2010}, 2400 \Delta_{7-2-2010}\}$$

### STEP 6

Comparing the relative change values in STEP 5 against the statistics in STEP 4, allows a conclusion on the index lagging or leading the historical performance.



### STEP 7

The data in STEP 6 is plotted in a various graphs to visualize the performance.

### RADIAL PLOTS

Each dataset  $S_i$  is plotted on a radial axis. All datasets are normalized to unit length.

$$i \delta_j = \frac{i \Delta_j - Min(\{S_i\})}{Max(\{S_i\}) - Min(\{S_i\})}$$

The mean, median and other statistics are plotted along the radial axis. The data from STEP 5 is similarly normalized and each interval is plotted along that axis.

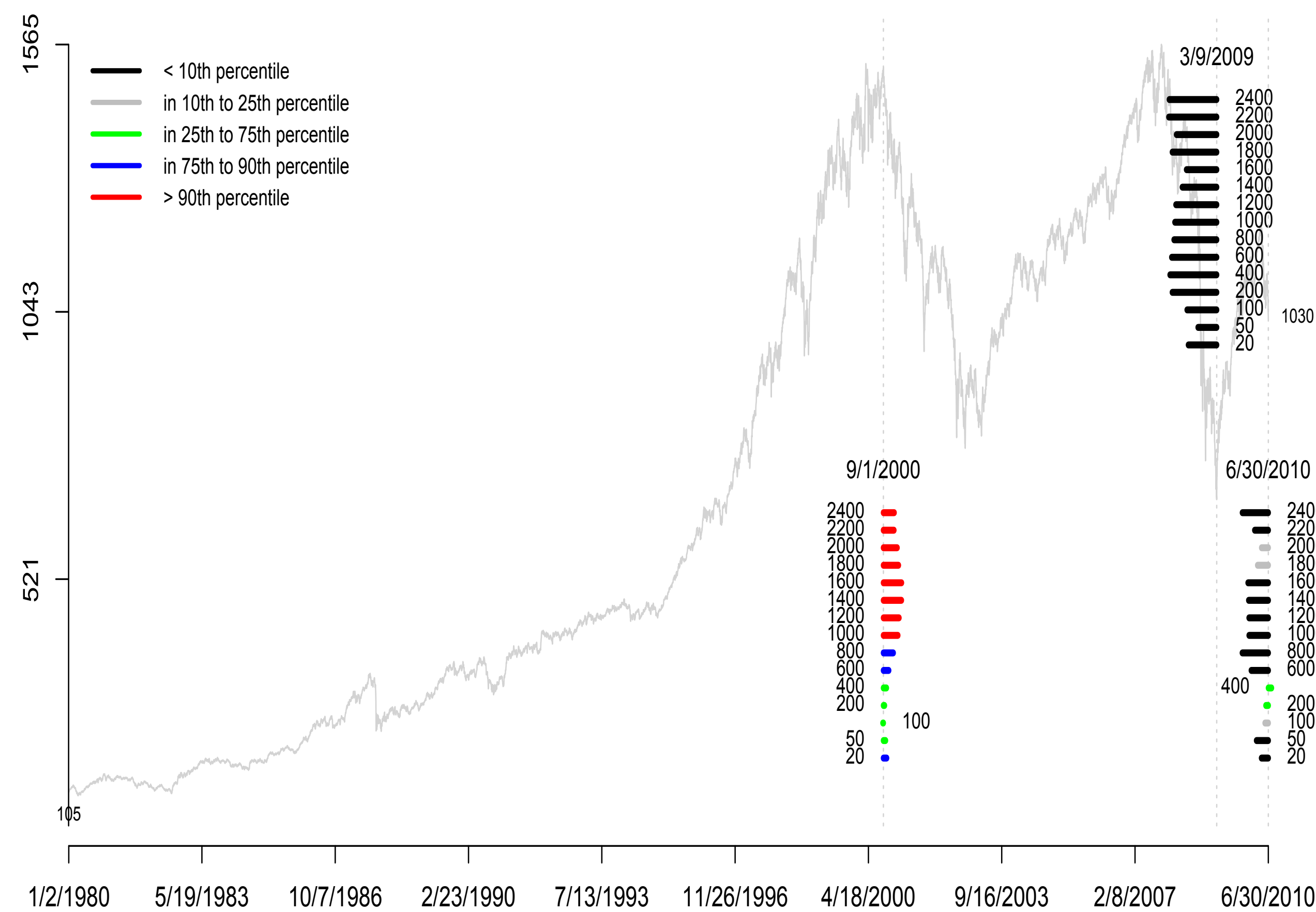
### HEAT MAPS

The entire dataset is plotted in horizontal plots, each individual plot representing one time interval. A vertical line of unit height and fixed width is plotted in 5 different colors. Green if it lies within the 25<sup>th</sup> and 75<sup>th</sup> percentiles, Blue if between 75<sup>th</sup> and 90<sup>th</sup> percentile, Grey if between 10<sup>th</sup> and 25<sup>th</sup> percentile, Black if less than the 10<sup>th</sup> percentile or Red if greater than the 90<sup>th</sup> percentile (i.e the outliers).

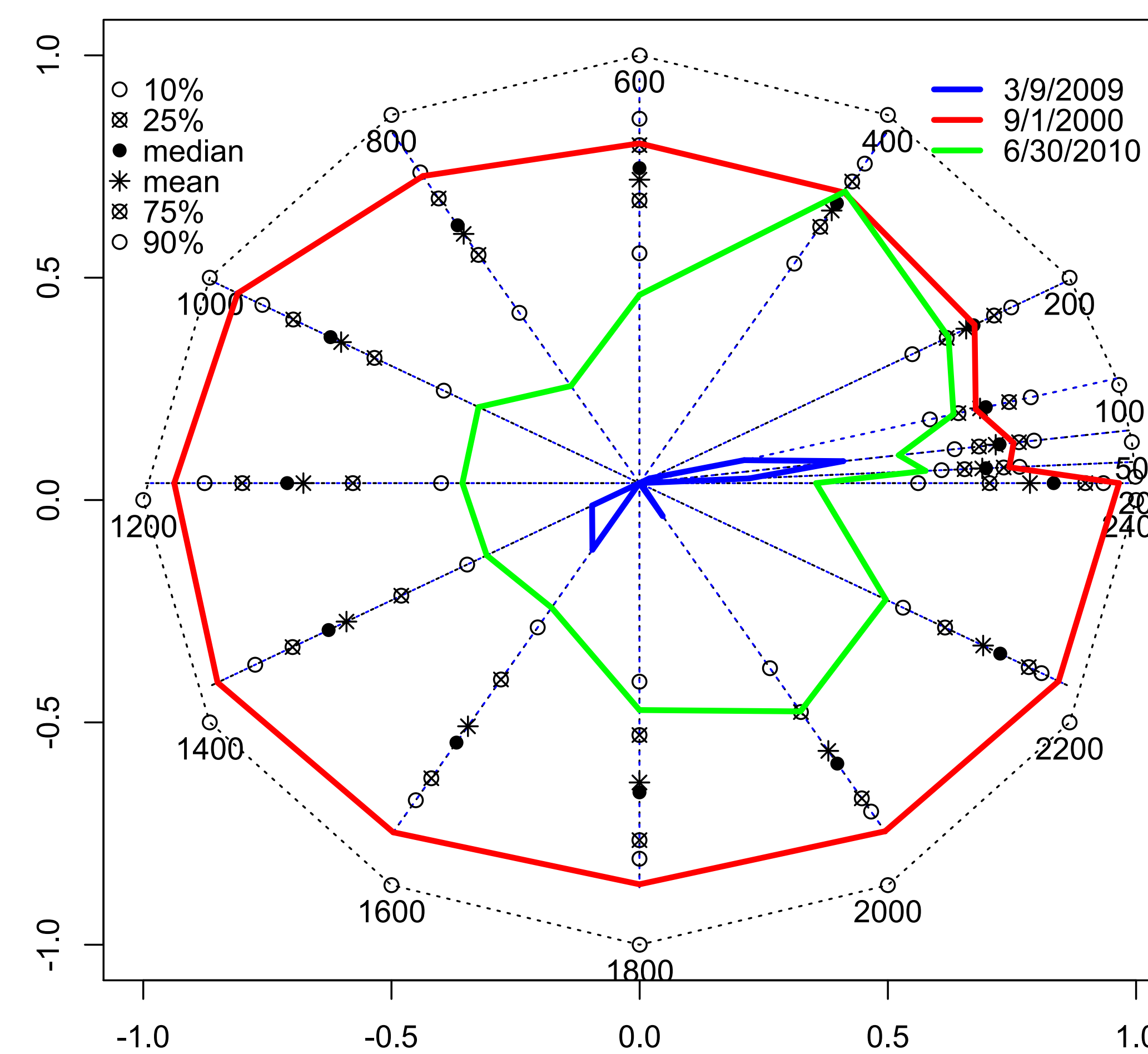
### RUN CHARTS

The entire normalized dataset value is plotted in 5 different colors chosen as in Heat Maps. Each plot represents one dataset, and plots are stacked vertically.

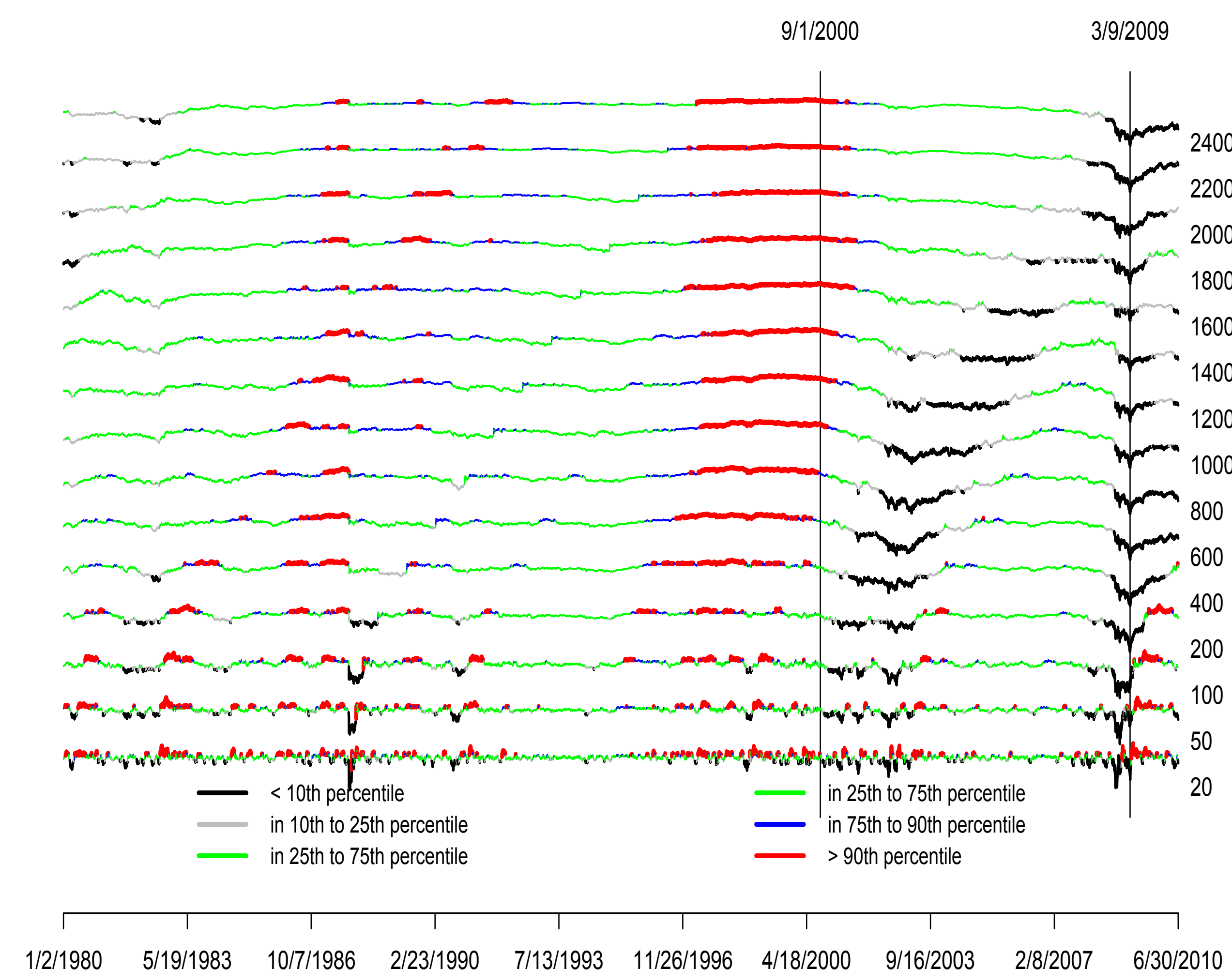
S&P 500 Cartesian Plot: Crash of 2009



S&P 500 Radial Plot: Crash of 2009



S&P 500 Run Chart: Crash of 2009



S&P 500 Heat Chart: Crash of 2009

