Eye-balling is OK Again: Using an Eye-Tracker to See How People Read Data Plots

Di Cook
with Yifan Zhao, Mahbub Majumder, Heike Hofmann, Omesh Johar, Niladri Roy Chowdhury
Two experiments

“Mind Reading: Using an Eye-Tracker to See How People are Looking at Lineups”
Zhao, Cook, Hofmann, Majumder, Roy Chowdhury (IJITAS 2013)

“The Statistical Eye: Reading Patterns in Visual Displays of Data”
Johar, Cook, Roy Chowdhury (2014, not published yet)
Looking at Lineups (Zhao et al)

12 lineups (plots examined)
24 subjects
In which plot, is the regression line the steepest?

- Example lineup - one has a significant slope, 19 generated from 0 slope model
- Easy one
- Expect to see clear reading pattern
• Example lineup
• Difficult
• Plots 6, 7 are very similarly strong slope
• Was this what people were looking at?
- All lineups, all subjects
- Lineups columns
- Subjects rows
- Visits to each plot and order of visit
Start --> Answer

Time looking at the lineup

Plot looked at, 1 ... 20

All subjects
• Subjects tended to go left to right, right to left in order
• Or from top to bottom, bottom to top
Experiment 2

- Johar et al, 2014
- Just look at ONE plot at a time
- How do people read particular statistical plots?
- What parts of the plot do they visit?
Three components

- Scatterplot of two associated variables: do eyes track the trend, see contamination?
- Side-by-side boxplots: read difference based on the median difference, box difference or outliers?
- Scatterplot of two variables containing clusters: gaps between clusters of points or color was the more dominant?
Experiment

- 36 subjects, recruited from Facebook
- Doctoral students from ISU, two undergrads
- Eye-movements were calibrated
- 8 scatter plots x2, a set of 9 box plots x2, and a set of 6 cluster plots x2
- Subjects had 3 seconds to look at plots
- Personality questionnaires administered
Does the plot margin affect detection of outliers? It has been reported that plots need to have a 5% margin between the most extreme data points and the axis for people to reasonably notice the extreme data elements. To test this, we plotted the same data using three different margins: 0%, 2.5%, and 5%. Larger than 5% creates too much whitespace. These levels were examined under different levels of association and whether or not a line was overlaid.

Sample of plots shown to subjects:

- **Margin=0**
- **Margin=2.5%**
- **Margin=5%**

<table>
<thead>
<tr>
<th>Cont=1%</th>
<th>Cont=5%</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Margin=0 Cont=1%" /></td>
<td><img src="image2" alt="Margin=0 Cont=5%" /></td>
</tr>
<tr>
<td><img src="image3" alt="Margin=2.5% Cont=1%" /></td>
<td><img src="image4" alt="Margin=2.5% Cont=5%" /></td>
</tr>
<tr>
<td><img src="image5" alt="Margin=5% Cont=1%" /></td>
<td><img src="image6" alt="Margin=5% Cont=5%" /></td>
</tr>
</tbody>
</table>
Factors

- Association: $r=0.5, 0.7, +/-$
- Regression line: present, absent
- Contamination: 1%, 5% (all four corners)
- Margin: 0, 2.5, 5%

Is the association between these two variables strong, moderate or weak?
Expectations

- Presence of line might draw eye to trend more
- Weaker association might mean people miss the contamination
- Smaller margin might mean people miss contamination, especially at 1%
DATA

- **fixation location** (x and y coordinates of the gaze point on the monitor screen; a gaze longer than 22 ms classifies as a fixation)
- **dwell time** (total time spent fixating, and not just glancing, on an area of interest)
- **glance count** (the number of times a participant looked at an area)

Plot fixations
Density of fixations on contamination increases with larger margin
We might expect a slightly higher variance in the eye positions if there wasn't a line association, slope of a regression line, outliers and plot margins.

Margin effect much stronger with 1% contamination.
To examine whether subjects noticed the contamination, the sum of the dwell time for each subject, relative to each factor, was computed. Focusing only on the dwell time in the outlier area of interest is summarized in the plots below. Dwell time is plotted against factor level. Only for contamination does it look like there is a difference in the amount of dwell time.

Not clear whether line helps focus on trend
Gap or color?
Does gap or color indicate clusters?

Healey & Enns (2011) report that color is more pre-attentively dominant than shape, when examining point clouds. In this experiment we are interested to know whether the gap between clusters or color of points is the dominant focus when people are asked to report on cluster separation. Four types of color application were made to the same plot of the data: no colouring, color that matches the gaps between clusters, color that splits the separated clusters, color that is just differently applied. We would expect when there is no color that people will naturally focus on the gap, but once color is applied that this might change.

Is gap or color dominant in perceiving clusters?

Cleveland & McGill (1984) report an ordering of elementary perceptual tasks related to reading information from data plots:

1. Position along a common scale
2. Position along unaligned scale
3. Length, direction, angle
4. Area
5. Volume, curvature
6. Color, shading

Below is a sample of the type of plots shown to subjects.

<table>
<thead>
<tr>
<th>Var=1</th>
<th>Var=5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color=none</td>
<td>Color=match</td>
</tr>
</tbody>
</table>

[Images of scatter plots for Var=1 and Var=5 with different color applications]
Factors

- Application of color: none, match, split, target
- Variance: 1, 3, 5

Is the separation between clusters a lot, some or a little?
Subjects' eyes visit the clusters more when color is applied.
When color is added that matches the clusters, people are drawn to look at all of the clusters.

When color is added that mismatches the clusters, people are drawn to the small target patch of color, and perhaps secondarily to the gaps.

It may be that when there is no color, that people primarily look at the gap, is hinted at when the densities are overlaid.

The eye movement varies a little when there is a larger variance in the data, with perhaps broader movement with more data spread.
THAT’S IT
This experiment investigated the effect of four factors:

- **Association**: correlation = 0.5, 0.7
- **Presence of regression line**: lm, none
- **Contamination**: 1%, 5%
- **Margin between data and axis**: 0%, 2.5%, 5%

In addition, we manipulated two other factors, namely the direction of association (positive or negative) and the location of outliers (top left or top right in the case of positive association and top right or bottom left in the case of negative association). Data were simulated and plots were made under all combinations of these conditions, yielding a total of 96 plots.

Thus, there were eight blocks denoting combinations of correlation, direction of association, and location of outliers. In each block there were twelve different combinations of presence of regression line, contamination, and margin.

This code reads in the data files for the linear component in order to report the design in tabular form.

Below are some summaries that show how the randomization was realized in relation to the treatments. We can see that all treatments were represented, but the design was not balanced.

### Source: local data frame [24 x 5]

<table>
<thead>
<tr>
<th>ln</th>
<th>marg</th>
<th>cont</th>
<th>cr</th>
<th>numsubjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>lm</td>
<td>0</td>
<td>0.01</td>
<td>0.5</td>
<td>12</td>
</tr>
<tr>
<td>lm</td>
<td>0</td>
<td>0.01</td>
<td>0.7</td>
<td>18</td>
</tr>
<tr>
<td>none</td>
<td>0.05</td>
<td>0.05</td>
<td>0.7</td>
<td>18</td>
</tr>
<tr>
<td>none</td>
<td>0.05</td>
<td>0.05</td>
<td>0.5</td>
<td>24</td>
</tr>
<tr>
<td>none</td>
<td>0.025</td>
<td>0.05</td>
<td>0.7</td>
<td>18</td>
</tr>
<tr>
<td>none</td>
<td>0.025</td>
<td>0.05</td>
<td>0.5</td>
<td>24</td>
</tr>
<tr>
<td>lm</td>
<td>0.05</td>
<td>0.01</td>
<td>0.7</td>
<td>15</td>
</tr>
<tr>
<td>lm</td>
<td>0.05</td>
<td>0.01</td>
<td>0.5</td>
<td>21</td>
</tr>
<tr>
<td>none</td>
<td>0.025</td>
<td>0.01</td>
<td>0.5</td>
<td>21</td>
</tr>
<tr>
<td>none</td>
<td>0.025</td>
<td>0.01</td>
<td>0.7</td>
<td>18</td>
</tr>
<tr>
<td>lm</td>
<td>0.05</td>
<td>0.05</td>
<td>0.7</td>
<td>24</td>
</tr>
<tr>
<td>lm</td>
<td>0.05</td>
<td>0.05</td>
<td>0.5</td>
<td>15</td>
</tr>
<tr>
<td>none</td>
<td>0.05</td>
<td>0.01</td>
<td>0.7</td>
<td>15</td>
</tr>
<tr>
<td>none</td>
<td>0.05</td>
<td>0.01</td>
<td>0.5</td>
<td>21</td>
</tr>
<tr>
<td>none</td>
<td>0</td>
<td>0.05</td>
<td>0.5</td>
<td>15</td>
</tr>
<tr>
<td>none</td>
<td>0</td>
<td>0.05</td>
<td>0.7</td>
<td>18</td>
</tr>
<tr>
<td>lm</td>
<td>0.025</td>
<td>0.05</td>
<td>0.5</td>
<td>21</td>
</tr>
<tr>
<td>lm</td>
<td>0.025</td>
<td>0.05</td>
<td>0.7</td>
<td>18</td>
</tr>
<tr>
<td>none</td>
<td>0</td>
<td>0.05</td>
<td>0.1</td>
<td>15</td>
</tr>
<tr>
<td>none</td>
<td>0</td>
<td>0.05</td>
<td>0.7</td>
<td>18</td>
</tr>
<tr>
<td>lm</td>
<td>0.025</td>
<td>0.01</td>
<td>0.5</td>
<td>21</td>
</tr>
<tr>
<td>lm</td>
<td>0.025</td>
<td>0.01</td>
<td>0.7</td>
<td>12</td>
</tr>
<tr>
<td>lm</td>
<td>0</td>
<td>0.05</td>
<td>0.5</td>
<td>9</td>
</tr>
<tr>
<td>lm</td>
<td>0</td>
<td>0.05</td>
<td>0.7</td>
<td>9</td>
</tr>
</tbody>
</table>

Sample sizes for each level not the same, but pretty well covered.
A linear mixed effects model is fit to the data. All factors were included initially, but only margin and contamination significantly predicted dwell time. Interaction between margin and correlation and presence of a regression line. Interaction between margin and correlation significantly predicted dwell time. A margin of 5% when correlation was 0.7 indicated a longer dwell time on the outlier. Contamination of 5% leads to a longer dwell time on the outlier.

## Linear mixed model fit by REML ['lmerMod']

Formula: totaldwelltime ~ marg * cr + cont + (1 | Subject)

Data: dwell.out

REML criterion at convergence: 4954

Random effects:

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>(Intercept)</td>
<td>187</td>
<td>13.7</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>8655</td>
<td>93.0</td>
</tr>
</tbody>
</table>

Number of obs: 420, groups: Subject, 36

Fixed effects:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>4.72</td>
<td>14.01</td>
<td>0.34</td>
</tr>
<tr>
<td>marg0.025</td>
<td>-3.69</td>
<td>16.71</td>
<td>-0.22</td>
</tr>
<tr>
<td>marg0.05</td>
<td>5.58</td>
<td>16.73</td>
<td>0.33</td>
</tr>
<tr>
<td>cr0.7</td>
<td>-18.71</td>
<td>17.61</td>
<td>-1.06</td>
</tr>
<tr>
<td>cont0.05</td>
<td>42.84</td>
<td>9.17</td>
<td>4.67</td>
</tr>
<tr>
<td>marg0.025:cr0.7</td>
<td>25.02</td>
<td>23.28</td>
<td>1.07</td>
</tr>
<tr>
<td>marg0.05:cr0.7</td>
<td>51.81</td>
<td>23.24</td>
<td>2.23</td>
</tr>
</tbody>
</table>

Correlation of Fixed Effects:

<table>
<thead>
<tr>
<th></th>
<th>(Intr)</th>
<th>marg0.025</th>
<th>marg0.05</th>
<th>cr0.7</th>
<th>cont0.05</th>
<th>marg0.025:cr0.7</th>
<th>marg0.05:cr0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>marg0.025</td>
<td>-0.733</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.759</td>
<td>-0.062</td>
</tr>
<tr>
<td>marg0.05</td>
<td>-0.735</td>
<td>0.615</td>
<td></td>
<td></td>
<td>-0.759</td>
<td>-0.062</td>
<td>-0.061</td>
</tr>
<tr>
<td>cr0.7</td>
<td>-0.705</td>
<td>0.584</td>
<td>0.584</td>
<td></td>
<td>0.584</td>
<td>0.584</td>
<td>0.576</td>
</tr>
<tr>
<td>cont0.05</td>
<td>-0.313</td>
<td>-0.002</td>
<td>-0.001</td>
<td>0.027</td>
<td>0.027</td>
<td>0.027</td>
<td>0.027</td>
</tr>
<tr>
<td>marg0.025:cr0.7</td>
<td>0.546</td>
<td>-0.718</td>
<td>-0.442</td>
<td>-0.759</td>
<td>-0.062</td>
<td>-0.718</td>
<td>-0.759</td>
</tr>
<tr>
<td>marg0.05:cr0.7</td>
<td>0.548</td>
<td>-0.441</td>
<td>-0.719</td>
<td>-0.759</td>
<td>-0.062</td>
<td>-0.441</td>
<td>-0.759</td>
</tr>
</tbody>
</table>

Participants typically noticed the contamination when the %contamination was higher, and when the margin was larger. The main effect of contamination level was significant, b = 42.84, t = 4.67, 95% CI [24.99, 60.73]. Moreover, a significant interaction between margin and correlation was observed, b=51.81, t = 2.23, 95% CI [6.44, 97.07].

### Summary

Association, slope of a regression line, outliers and plot margins file:///C:/Users/omesh/Dropbox/Omesh-CC/experiment/analysis/linear.html

Mixed effects model indicates dwell time on contamination affected by %contamination, and the larger margin in relation to stronger association
Reading side-by-side boxplots
Stimuli
Each participant saw eighteen plots. Every plot came from a different block and was randomly selected from the three plots in the corresponding block. As mentioned earlier, a specific block denoted a particular combination of median difference, block difference, and outlier. The randomly selected plots were then split into two subsets. Thus, we generated twelve modules of eighteen plots (two sets of nine). Participants were randomly assigned to one of these modules.

Apparatus
Plots were loaded into a 500 Hz binocular remote eye-tracker (SMI RED500). The eye-tracker randomized the order of plots within each of the two sets of nine. The Experiment Center software was used to display the plots. The BeGaze software was used to export eyetracking data. The software recorded the position of the gaze on the screen indexed by time.

Procedure
Participants were told that they would look at different kinds of box plots and will be asked simple questions about each plot. They were told that they had 3 seconds to look at a plot, following which they would be asked, “By how much is group B bigger than group A: a lot, some, or none?” Therefore, before participants looked at a plot they were aware of the question to follow. The rationale for giving this task was that we are interested in what people look at when asked to compare two samples using boxplots.
Factors

Median difference: 0, 1, 2
Box difference: 1, 1.5, 2
Outliers: yes, no

By how much is group B bigger than group A, a lot, some or none?
Results

BeGaze software provided the fixation location (x and y coordinates of the gaze point on the monitor screen; a gaze longer than 22 ms classifies as a fixation) indexed by time. The plots below show the fixation locations, with eye positions connected by time of visit. The layout of the plots follows median difference in the columns and box size difference in the rows. A loess smoother is overlaid on the points to give a quick sense of the direction of the eye movements. The first set of plots shows the data for the plots without outliers, and the second set shows those with outliers included. The last set of plots shows the loess smoothers for outliers present or absent.
The rough patterns of viewing follow similarly to expected. As the median increases people are clearly looking at the two median lines, as seen from the increase eye positions from left to right as...
Pairs of color application are compared: none and matching
Pairs of color application are compared: match vs target
Pairs of color application are compared: match vs split

The main findings from this component of the experiment are:

- When color is added that matches the clusters, people are drawn to look at all of the clusters.
- When color is added that mismatches the clusters, people are drawn to the small target patch of color, and perhaps secondarily to the gaps.

It may be that when there is no color, that people primarily look at the gap, as hinted at when the densities are overlaid.

The eye movement varies a little when there is a larger variance in the data, with perhaps broader movement with more data spread.

To fully assess the effect of color additional processing of the data is needed. We need to compute specifically whether the eyes track into the gaps between clusters, and whether this happens more without color. It would be interesting to note the first feature viewed by each person, and whether this is different under the different styles of color application.
Findings

- The amount of contamination has the biggest effect on whether people notice outliers, and a margin of 5% gives the best chance of discovery, particularly when association is strong.

- People take more notice of the colors than the gaps when clusters are colored, and tend to visit all of the clusters more thoroughly. When color does not match the clusters, when a small target area is coloured, people are drawn to this, and perhaps could neglect the gaps between clusters.
In reading boxplots to compare two samples, people do compare the medians, as well as the size of the box. Outliers don't interfere with this.
To Do

- More data cleaning
- Match subject answers with eye patterns
- Look at dwell location and time with the cluster/color experiment
- Examine personality assessments with responses