

FIELD OF INQUIRY

Data graphics visually display measured quantities by means of the combined use of points, lines, a coordinate system, numbers, symbols, words, shading and colour.

2,500,000,000,000,000,000

Data → Familiarize → Visualize → Analyze and Explore → Insight

– Richard Saul Wurman



TRY GLOBE UNLIMITED - 1 MONTH FOR JUST 99¢



And get unlimited access on all your devices

See my options!



Civic leaders urge Toronto mayor to speak up on drug allegations

- **OPINION** Rob Ford, Charbonneau and the dark side of social media
- **GLOBE EDITORIAL** Rob Ford's friends need to demand that he do the right thing
- **WATCH** Video: Ontario premier speaks out on Rob Ford crack allegations

EDITORS' PICKS



COMMENTARY



ARTS



GLOBE INVESTOR



REPORT ON BUSINESS

Canada drops out of top 20 in global science ranking 46

OPINION NDP's Horwath stumbles in her grand gesture to support Ontario budget 28

SNC granted special status in fraud case against former executive 0

Lack of interpreters in Ontario courts becoming intolerable, judge says 25

Duhatschek: A tale of two Thorntons 0

Rogers' Zoocasa real-estate brokerage reduces agents' commissions 0

Why Ben Bernanke will determine whether you have a good or bad day 0

Slowdown? Nearly half of Canadian home owners eager to buy property 22

INTRODUCING THE ALL-NEW
2013 LINCOLN MKZ

DISCOVER YOURS

Legal | THE LINCOLN MOTOR COMPANY

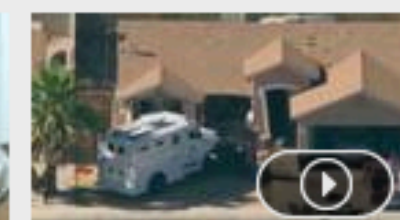
VIDEO »



NEWS
Jodi Arias says death penalty would cause her family more pain



REPORT ON BUSINESS
Microsoft unveils Xbox One



INITIATION

Visualizing Globe and Mail's Demographics and Business data sets with 1023 rows and 110 and 80 columns of data, respectively.

*Exploring and Explaining Data with
the Processing Environment*

Visualizing Data



O'REILLY®

Ben Fry

FRAMEWORK

Acquire

Obtain the data, whether from a file on a disk or a source over a network.

Parse

Provide some structure for the data's meaning, and order it into categories.

Filter

Remove all but the data of interest.

Mine

Apply methods from statistics or data mining as a way to discern patterns or place the data in mathematical context.

Represent

Choose a basic visual model, such as a bar graph, list, or tree.

Refine

Improve the basic representation to make it clearer and more visually engaging.

Interact

Add methods for manipulating the data or controlling what features are visible.

– Ben Fry's seven-step data visualization framework

Literature Review

VISUAL PERCEPTION

A three-stage model:

- 1- Simultaneous extraction of basic features*
- 2- Serial processing of patterns and structures*
- 3- Goal-oriented reduction of information*

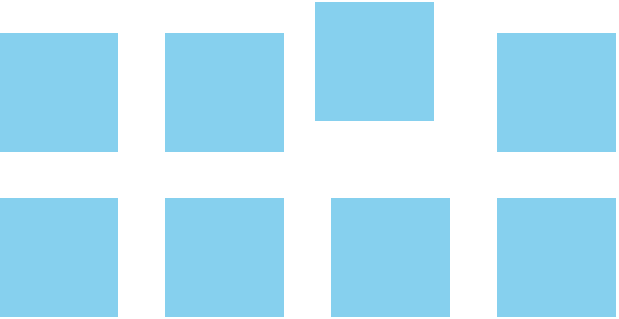
PRE-ATTENTIVE VARIABLES

Designers can make intentional use of pre-attentive variables in stage one to increase the performance of target detection, boundary detection, region tracking, counting and estimation.

PRE-ATTENTIVE VARIABLES



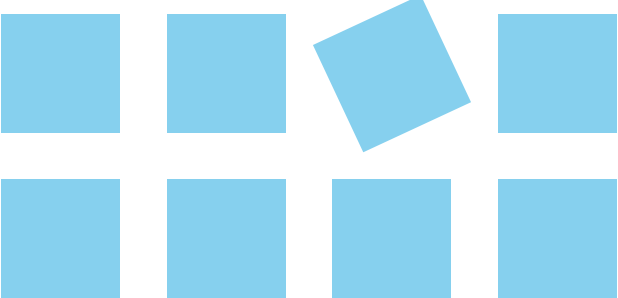
Colour Hue



Position and Alignment



Colour Brightness



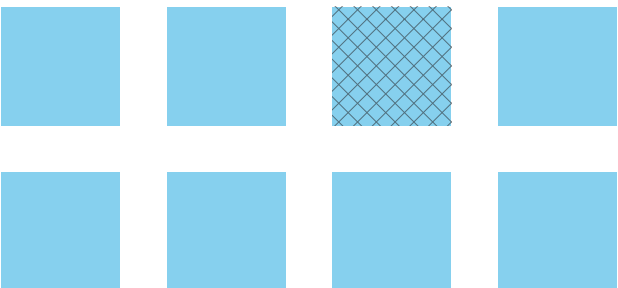
Orientation



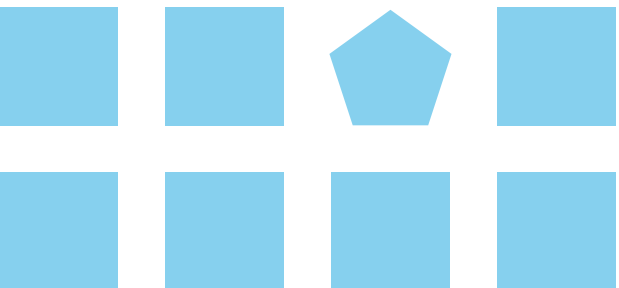
Colour Saturation



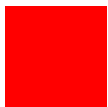
Size

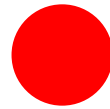


Texture



Shape







PRE-ATTENTIVE VARIABLES

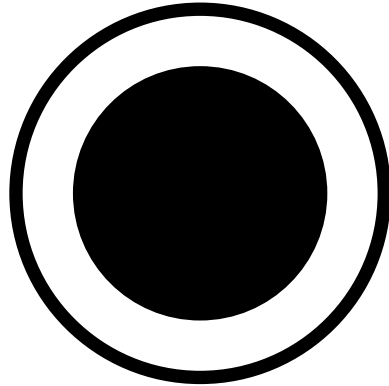
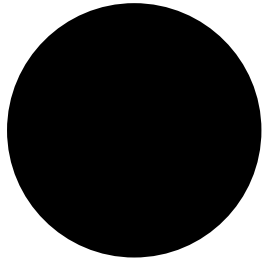
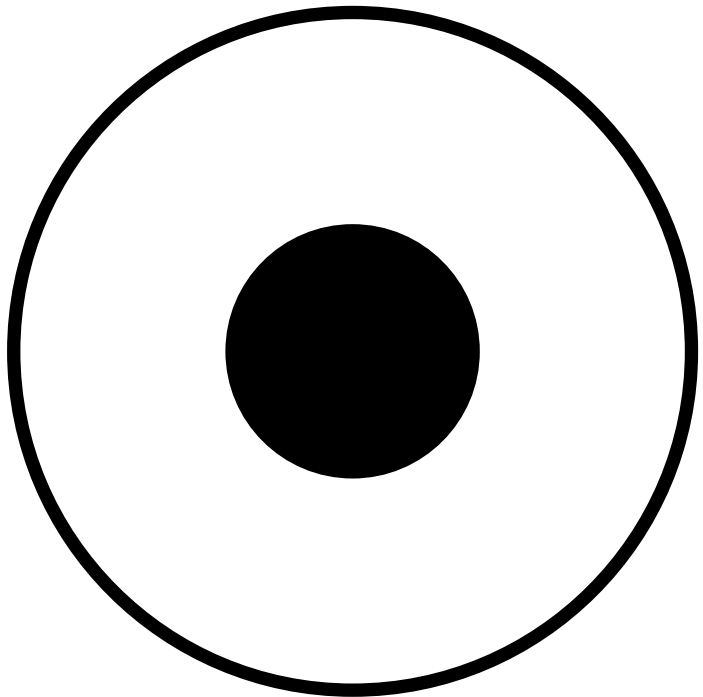
However there are factors that might impair the detection of pre-attentive-designed elements, such as number and variety—the degree of differentiation—of distractors in the representation.

RETINAL LEGIBILITY

In order to achieve maximum differentiation, it is necessary: (1) to have a total amount of size sufficient so that the smallest signs are visible, stand-out from the background and cover the visual noise, but limited so that the largest signs do not overlap and are separate (2) to obtain the greatest amount of differentiation , that is, to utilize the entire perceptible range of a given variable.

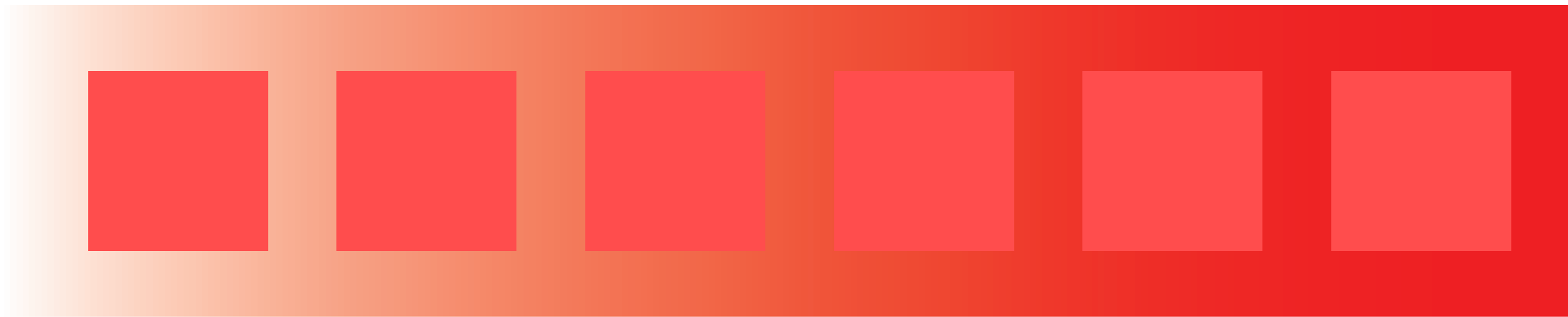
DELBOEUF ILLUSION

The Delboeuf illusion is an optical illusion of relative size perception. In the best-known version of the illusion, two circles of identical size have been placed near to each other and one is surrounded by an annulus; the surrounded circle then appears larger than the non-surrounded circle if the annulus is close, while appearing smaller than the non-surrounded circle if the annulus is distant.

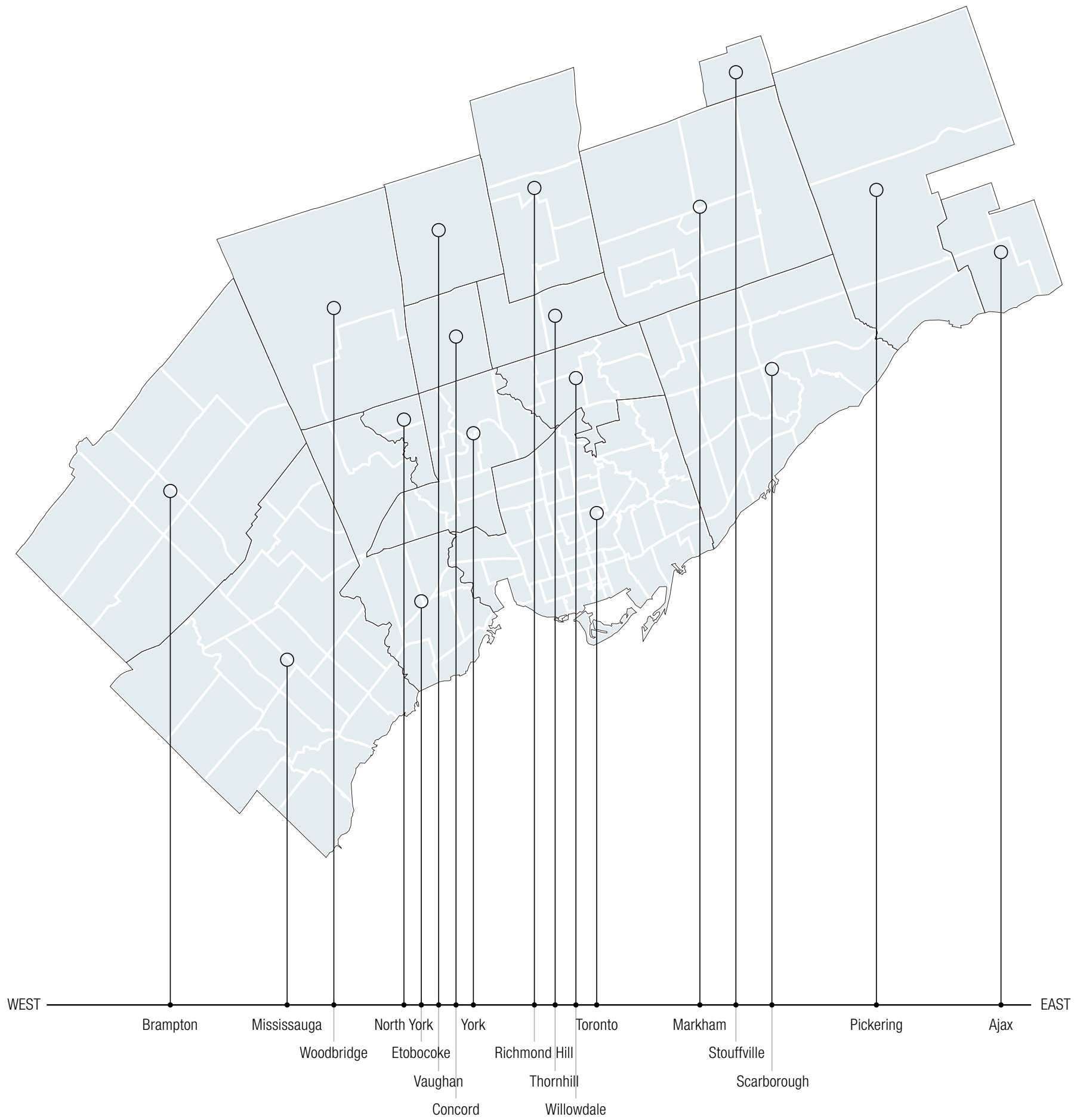


SIMULTANEOUS CONTRAST

Simultaneous contrast is an illusion that occurs when different hues, values or intensities, lie beside one another. Therefore colours that are not particularly strong when viewed by themselves might look brilliant when placed beside other colours.



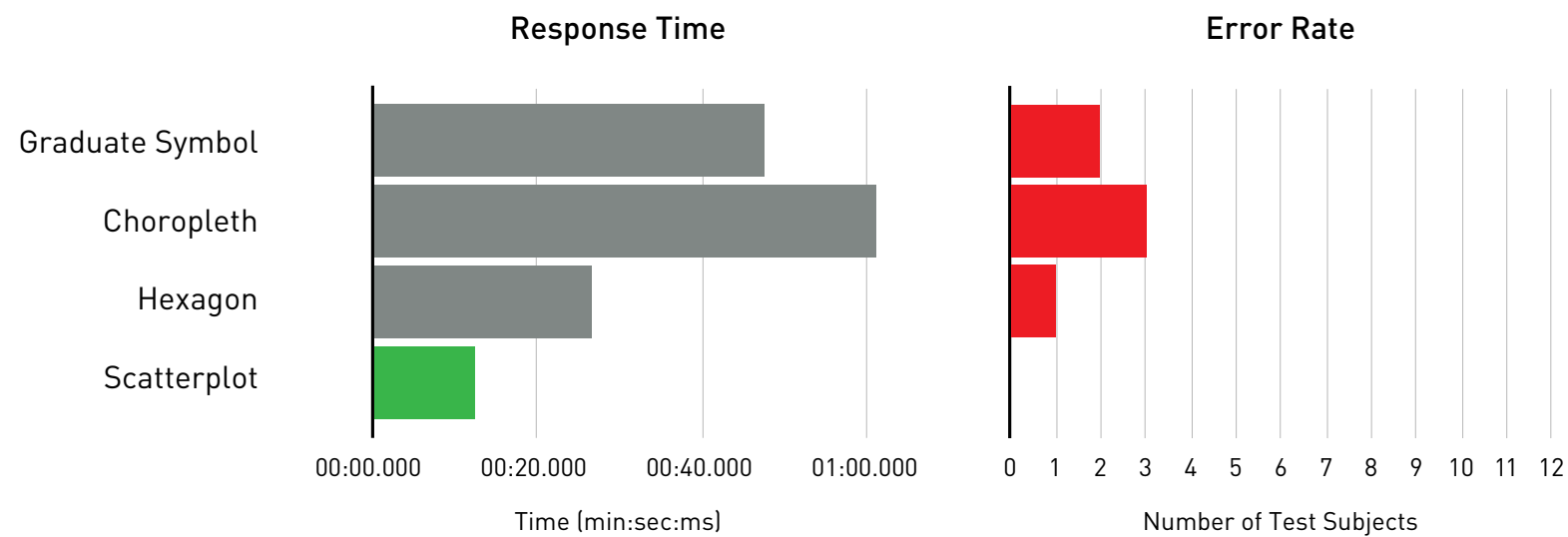
Prototypes



Evaluation

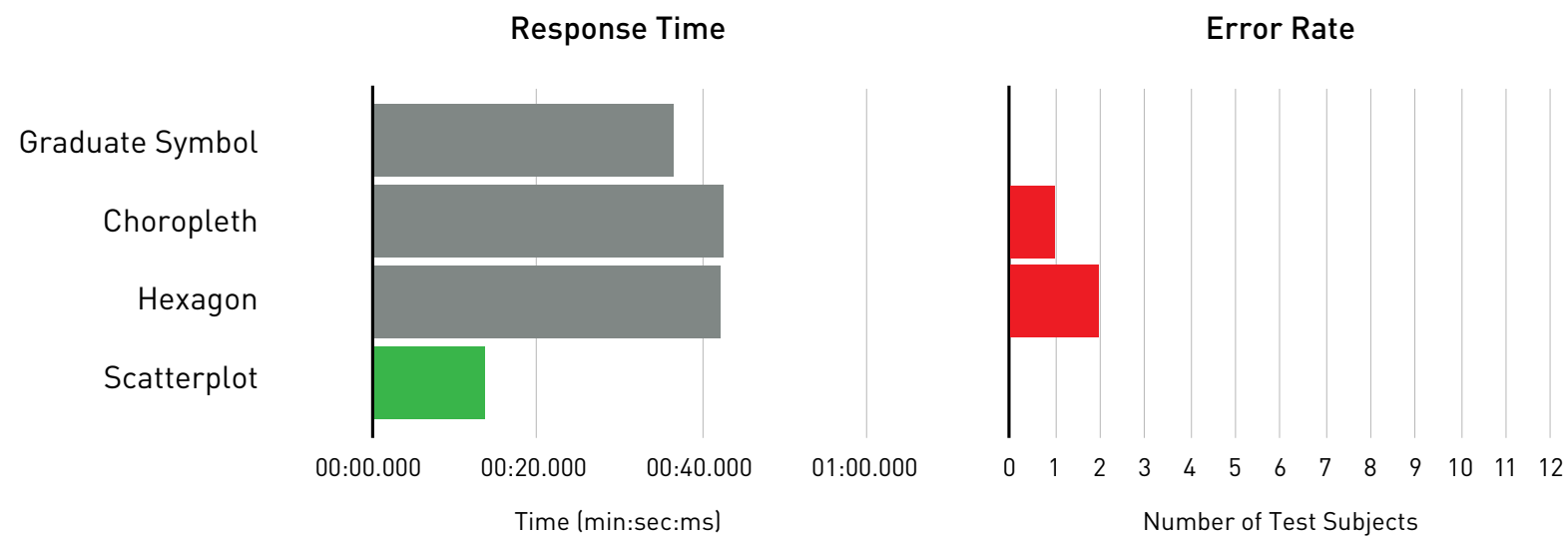
Quantitative Analysis

Order	Graduate Symbol (A)	Choropleth (B)	Hexagon (C)	Scatterplot (D)
CBAD	00:42.750	00:45.900	00:22.630	00:07.530
DCBA	01:31.190	01:24.650	00:25.530	00:15.380
BCDA	00:42.030	02:44.630	00:45.510	00:24.260
DACB	00:28.700	00:26.250	00:19.350	00:14.280
ACBD	01:45.310	01:38.730	00:33.050	00:11.360
CADB	01:03.350	00:50.520	00:49.900	00:11.110
CDAB	00:30.260	01:16.570	00:31.270	00:10.550
ADCB	00:36.130	00:37.170	00:13.380	00:12.650
BDAC	00:34.510	01:02.430	00:14.430	00:11.480
BADC	00:32.050	00:44.250	00:11.010	00:07.650
CABD	00:37.030	00:18.330	00:33.560	00:12.850
CDBA	00:24.800	00:33.780	00:20.480	00:12.910
Average	00:47.343	01:01.934	00:26.675	00:12.668



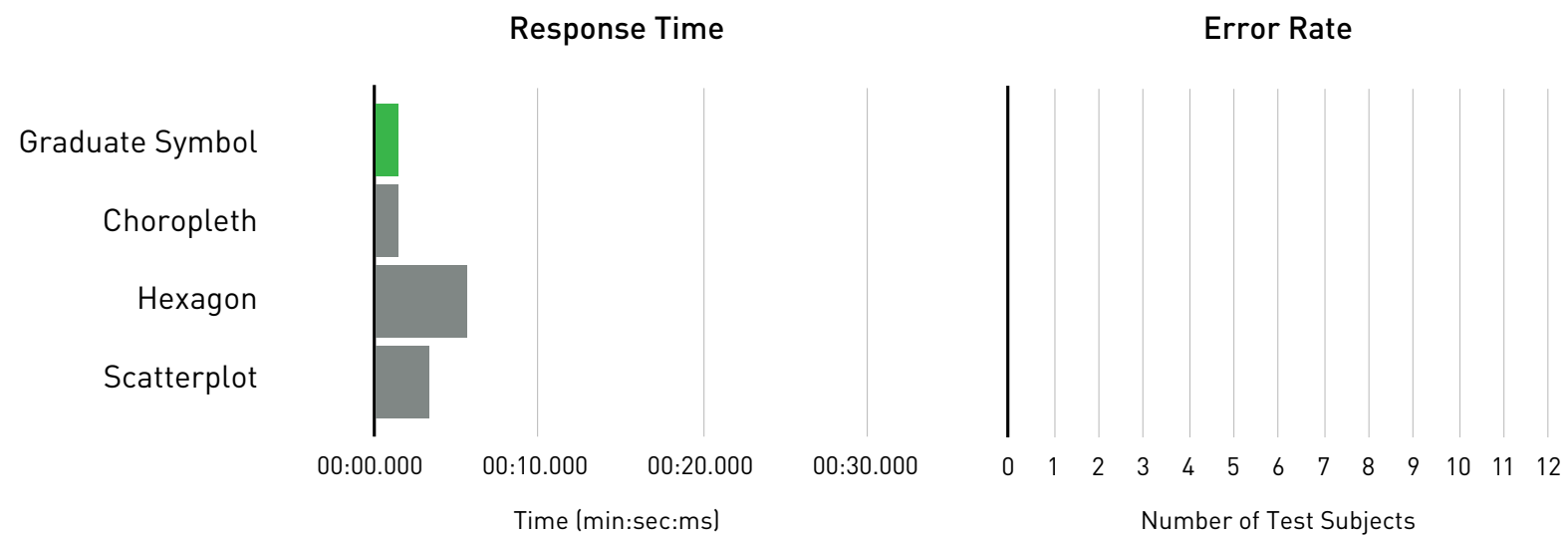
– User test results for finding and registering top three data points with highest values.

Order	Graduate Symbol (A)	Choropleth (B)	Hexagon (C)	Scatterplot (D)
CBAD	00:16.330	00:42.870	01:08.600	00:09.140
DCBA	01:11.220	01:04.580	01:03.030	00:20.660
BCDA	01:16.130	01:35.720	00:45.380	00:21.560
DACB	00:37.350	00:20.450	00:27.080	00:13.360
ACBD	01:04.280	01:01.300	00:56.170	00:10.650
CADB	00:39.470	00:25.180	01:18.140	00:12.150
CDAB	00:17.780	00:36.880	00:45.350	00:13.760
ADCB	00:32.100	00:32.060	00:19.730	00:15.000
BDAC	00:17.610	00:53.350	00:16.150	00:13.200
BADC	00:17.030	00:51.020	00:13.530	00:11.000
CABD	00:25.130	00:10.300	00:48.420	00:12.200
CDBA	00:21.830	00:19.130	00:23.180	00:13.510
Average	00:36.355	00:42.737	00:42.063	00:13.849



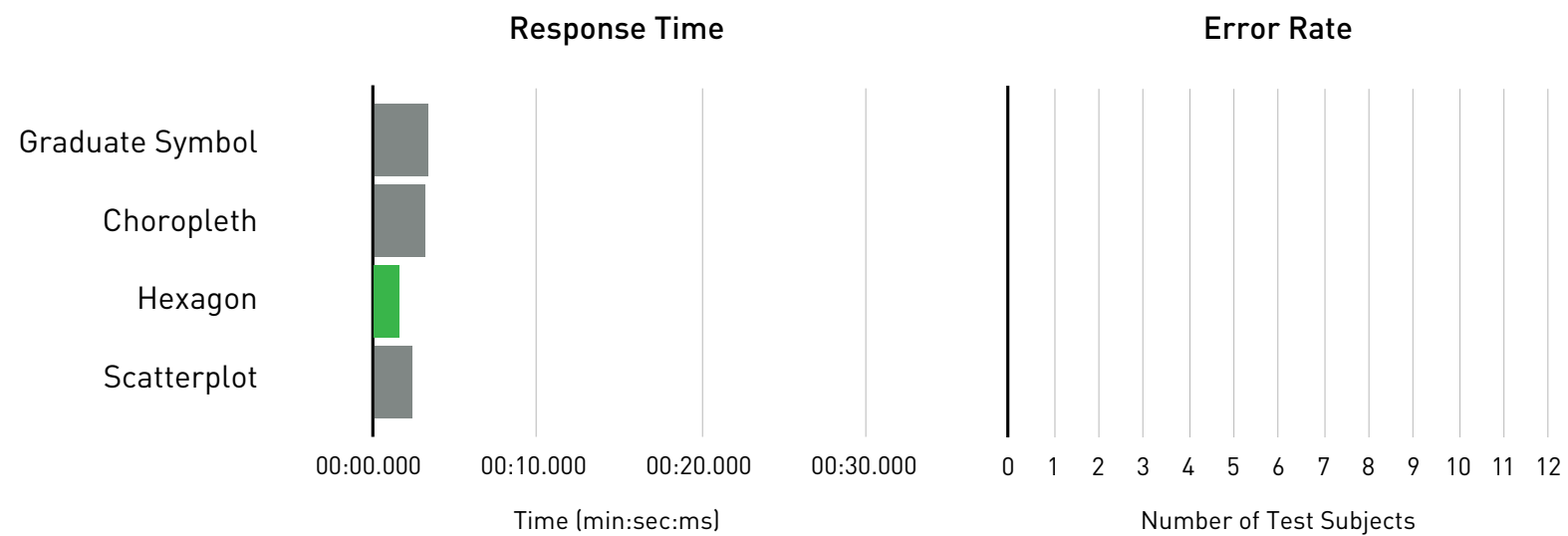
– User test results for finding and registering bottom three data points with lowest values.

Order	Graduate Symbol (A)	Choropleth (B)	Hexagon (C)	Scatterplot (D)
CBAD	00:00.500	00:01.990	00:04.030	00:02.400
DCBA	00:01.880	00:01.900	00:02.940	00:03.560
BCDA	00:00.160	00:01.190	00:04.460	00:02.960
DACB	00:02.900	00:00.780	00:04.160	00:02.810
ACBD	00:03.160	00:02.860	00:21.480	00:01.240
CADB	00:01.010	00:01.000	00:11.680	00:01.450
CDAB	00:00.410	00:00.330	00:03.510	00:00.180
ADCB	00:02.060	00:00.350	00:01.480	00:01.610
BDAC	00:01.560	00:04.010	00:03.800	00:08.110
BADC	00:01.110	00:01.360	00:00.960	00:01.600
CABD	00:01.410	00:01.350	00:05.410	00:11.980
CDBA	00:00.910	00:00.860	00:03.440	00:04.110
Average	00:01.423	00:01.498	00:05.612	00:03.501



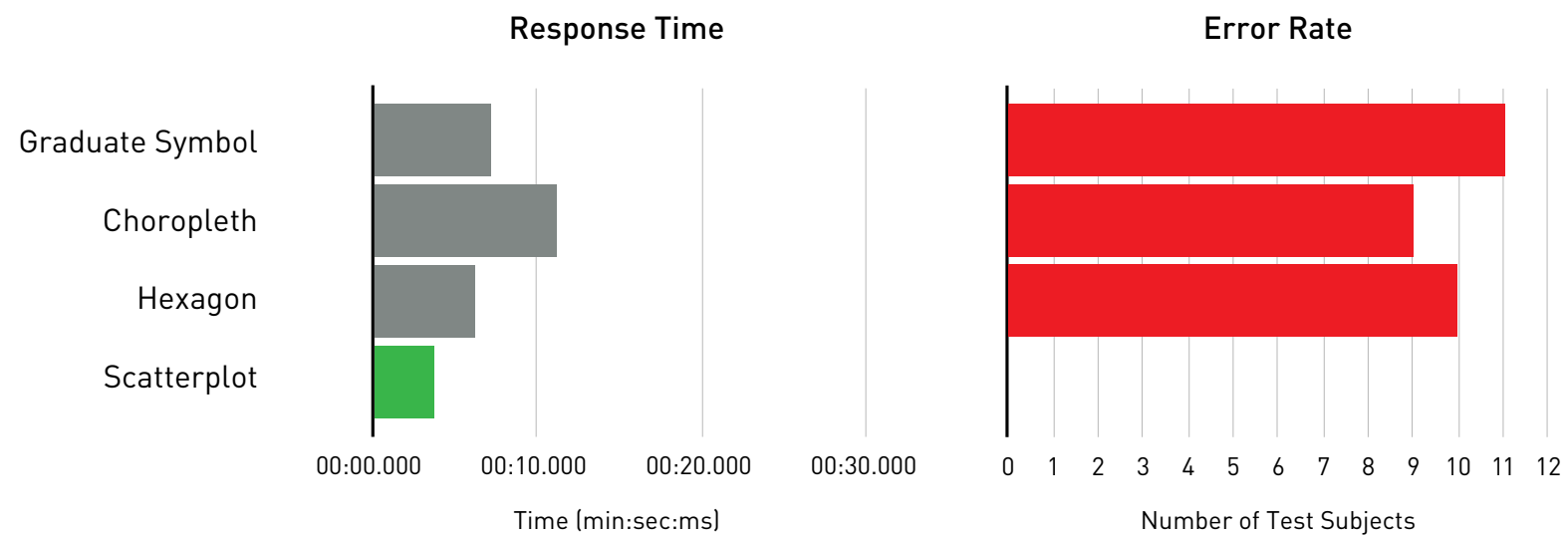
– *User test results for recognizing a region with clustering of data points with the highest values.*

Order	Graduate Symbol (A)	Choropleth (B)	Hexagon (C)	Scatterplot (D)
CBAD	00:00.400	00:00.680	00:02.240	00:00.300
DCBA	00:22.760	00:02.830	00:00.480	00:09.280
BCDA	00:00.360	00:17.000	00:01.010	00:00.350
DACB	00:02.900	00:01.510	00:00.210	00:09.910
ACBD	00:03.730	00:08.730	00:02.250	00:00.430
CADB	00:01.150	00:00.780	00:01.880	00:00.930
CDAB	00:00.400	00:01.000	00:07.350	00:00.200
ADCB	00:02.060	00:00.210	00:00.350	00:00.890
BDAC	00:00.510	00:03.500	00:00.460	00:00.810
BADC	00:00.650	00:01.330	00:00.530	00:00.280
CABD	00:03.730	00:00.580	00:01.050	00:03.550
CDBA	00:02.430	00:00.550	00:01.690	00:02.660
Average	00:03.423	00:03.225	00:01.625	00:02.466



– User test results for recognizing a region with clustering of data points with the lowest values.

Order	Graduate Symbol (A)	Choropleth (B)	Hexagon (C)	Scatterplot (D)
CBAD	00:04.280	00:30.530	00:07.350	00:10.110
DCBA	00:09.400	00:02.330	00:05.380	00:04.360
BCDA	00:09.800	00:09.580	00:01.380	00:00.510
DACB	00:00.480	00:03.160	00:05.160	00:00.760
ACBD	00:08.080	00:15.660	00:05.710	00:03.330
CADB	00:04.260	00:01.660	00:04.660	00:05.580
CDAB	00:01.330	00:00.380	00:06.550	00:04.040
ADCB	00:05.480	00:06.360	00:04.000	00:00.330
BDAC	00:17.130	00:38.410	00:16.960	00:04.550
BADC	00:09.110	00:05.230	00:05.750	00:10.460
CABD	00:09.510	00:19.000	00:02.830	00:00.410
CDBA	00:07.760	00:03.280	00:08.680	00:00.760
Average	00:07.218	00:11.298	00:06.201	00:03.767



– *User test results for comparing a data point of interest with remainder of the data points.*

Qualitative Analysis

CONCLUSIONS

What this study demonstrated was that the choice of a model should support the task at hand. Although map-based visualizations might be well suited for geographic analysis and spatial navigation, they are not also optimal for representing the actual values of data points.