3.5 Case Study: Purple America
Challenge. Visualize election results.

“If I can't picture it, I can't understand it.”
— Albert Einstein
Modular Programming

Modular programming.

- Model problem by decomposing into components.
- Develop data type for each component.

Polygon. Geometric primitive.
Region. Name, postal abbreviation, polygonal boundary.
Vote tally. Number of votes for each candidate.
Election map. Regions and corresponding vote tallies for a given election.

```
<table>
<thead>
<tr>
<th>ElectionMap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
</tr>
<tr>
<td>Region</td>
</tr>
<tr>
<td>VoteTally</td>
</tr>
<tr>
<td>VoteTally</td>
</tr>
<tr>
<td>String</td>
</tr>
<tr>
<td>String</td>
</tr>
<tr>
<td>Polygon</td>
</tr>
<tr>
<td>int</td>
</tr>
<tr>
<td>int</td>
</tr>
<tr>
<td>int</td>
</tr>
<tr>
<td>double[]</td>
</tr>
<tr>
<td>double[]</td>
</tr>
<tr>
<td>int</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
```

hierarchy of data types
Data Sources
Boundary Data: States within the Continental US

**Geometric data.** [US census bureau]
- www.census.gov/tiger/boundary
- NJ.txt has boundaries of every county in New Jersey.
- USA.txt that has boundary of every state.

**Election results.** [David Leip]
- http://uselectionatlas.org/RESULTS
- Interactive and graphical.
- Need to screen-scrape to get data.

**Emerging standard.**
- Publish data in text form on the web (like geometric data).
- Write programs to produce visuals (like we’re doing!)
- Mashups.
Boundary Data: States within the Continental US

USA data file. State names and boundary points.

% more USA.txt

-124.731216 24.544102 -66.980385 49.384365

104 number of regions

Alabama
USA
498
-88.200027 34.995548
-88.202919 35.007942

... New Jersey
USA
368
-74.695305 41.357330
-74.461754 41.250000
-74.366302 41.202801
... -74.721313 41.347294
...
**Boundary Data: Counties within a State**

State data files. County names and boundary points.

### % more NJ.txt

<table>
<thead>
<tr>
<th>Region</th>
<th>State</th>
<th>Code</th>
<th>Number of Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlantic</td>
<td>NJ</td>
<td>127</td>
<td>88</td>
</tr>
<tr>
<td>-74.877563</td>
<td>39.608414</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-74.736694</td>
<td>39.729721</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mercer</td>
<td>NJ</td>
<td>88</td>
<td>88</td>
</tr>
<tr>
<td>-74.748825</td>
<td>40.424248</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-74.722702</td>
<td>40.375301</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-74.674507</td>
<td>40.384399</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(comment: number of regions)

88 points

(longitude, latitude)

(comment: bounding box)

(comment: (-75.56, 38.92) (-73.89, 41.35))
Pitfalls: Pieces and Holes

**Pieces.** A state can be comprised of several disjoint polygons.

**Holes.** A county can be entirely inside another county.
Screen Scraping the Election Returns

Screen scrape. Download html from web and parse.

county name is text between <b> and </b> tags that occurs after width:100px

Election Scraper (sketch)

```java
int year    = 2008;    // election year
String usps = "NJ";    // United States postal code for New Jersey
int fips    = 34;      // FIPS code for New Jersey
String url  = "http://uselectionatlas.org/RESULTS/datagraph.php";
In in       = new In(url + "?year=" + year + "&fips=" + fips);
Out file   = new Out(usps + year + ".txt");
String input = in.readAll();

while (true) {
    // screen scrape county name
    int p = input.indexOf("width:100px", p);
    if (p == -1) break;
    int from = input.indexOf("<b>", p);
    int to   = input.indexOf("</b>", from);
    String county = input.substring(from + 3, to);

    // screen scrape vote totals for each candidate

    // save results to file
    file.println(county + "," + mccain + "," + obama + "," + other + ",");
}
```

extract text between `<b>` and `</b>` tags, that occurs after `width:100px`
More Pitfalls

Data sources have different conventions.
- FIPS codes: NJ vs. 34.
- County names: LaSalle vs. La Salle, Kings County vs. Brooklyn.

Plenty of other minor annoyances.
- Unreported results.
- Third-party candidates.
- Changes in county boundaries.

Bottom line. Need to clean up data (but write a program to do it!)
Polygons and Regions
Polygon Data Type

**Polygon.** Closed, planar path with straight line segments.

**Simple polygon.** No crossing lines.
public class Polygon {
    private final int N; // number of boundary points
    private final double[] x, y; // the points (x[i], y[i])

    // read from input stream
    public Polygon(In in) {
        N = in.readInt();
        x = new double[N];
        y = new double[N];
        for (int i = 0; i < N; i++) {
            x[i] = in.readDouble();
            y[i] = in.readDouble();
        }
    }

    public void fill() { StdDraw.filledPolygon(x, y); }

    public boolean contains(double x0, double y0) { ... }
    public String toString() { ... }
}
**Region Data Type**

**Region.** Represents a state or county.

Mercer, NJ
88 point polygon

New Jersey, USA
368 point polygon
public class Region {
    private final String name;  // name of region
    private final String usps;  // postal abbreviation
    private final Polygon poly; // polygonal boundary

    public Region(String name, String usps, Polygon poly) {
        this.name = name;
        this.usps = usps;
        this.poly = poly;
    }

    public void draw() { poly.fill(); }

    public boolean contains(double x0, double y0) {
        return poly.contains(x0, y0);
    }

    public String toString() { ... }
}

Election Returns
Election Returns: By State

Screen-scraping results. Votes for McCain, Obama, Other by region.

% more USA2008.txt
Alabama,1266546,813479,19773,
Alaska,193841,123594,8762,
Arizona,1230111,1034707,39020,
Arkansas,638017,422310,26290,
California,5011781,8274473,289260,
Colorado,1073584,1288568,39197,
Connecticut,629428,997772,19592,
Delaware,152374,255459,4579,
District of Columbia,17367,245800,2686,
Florida,4045624,4282074,82621,
Georgia,2048744,1844137,39222,
Hawaii,120566,325871,7131,
Idaho,403012,236440,17978,
Illinois,2031527,3419673,71851,
... Virginia,1725005,1959532,38723,
Washington,1229216,1750848,68820,
West Virginia,398061,304127,12550,
Wisconsin,1262393,1677211,43813,
Wyoming,164958,82868,6832,
Election Returns: By County

Screen-scraping results. Votes for McCain, Obama, Other by region.

<table>
<thead>
<tr>
<th>County</th>
<th>Votes for McCain</th>
<th>Votes for Obama</th>
<th>Votes for Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlantic</td>
<td>67830,1517</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bergen</td>
<td>225367,4424</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burlington</td>
<td>131219,2930</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Camden</td>
<td>159259,3304</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cape May</td>
<td>22893,802</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cumberland</td>
<td>34919,915</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Essex</td>
<td>240306,2181</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gloucester</td>
<td>77267,1848</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hudson</td>
<td>154140,2116</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hunterdon</td>
<td>29776,1147</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mercer</td>
<td>107926,2229</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middlesex</td>
<td>193812,4283</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monmouth</td>
<td>148737,4244</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morris</td>
<td>112275,2913</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ocean</td>
<td>110189,4111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passaic</td>
<td>113257,1904</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salem</td>
<td>112044,672</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Somerset</td>
<td>79321,1672</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sussex</td>
<td>28840,1393</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Union</td>
<td>28840,1393</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warren</td>
<td>20628,980</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Vote Tally Data Type

**VoteTally.** Represents the election returns for one region.

**Mercer, NJ**
- 50,223 McCain
- 107,926 Obama
- 2,229 Other

**New Jersey, USA**
- 1,613,207 McCain
- 2,215,422 Obama
- 47,826 Other
public class VoteTally {
    private final int rep, dem, ind;

    public VoteTally(String name, String usps, int year) {
        In in = new In(usps + year + " .txt");
        String input = in.readAll();
        int i0 = input.indexOf(name);
        int i1 = input.indexOf("", i0+1);
        int i2 = input.indexOf("", i1+1);
        int i3 = input.indexOf("", i2+1);
        int i4 = input.indexOf("", i3+1);
        rep = Integer.parseInt(input.substring(i1+1, i2));
        dem = Integer.parseInt(input.substring(i2+1, i3));
        ind = Integer.parseInt(input.substring(i3+1, i4));
    }

    public Color getColor() {
        if (rep > dem) return StdDraw.RED;
        if (dem > rep) return StdDraw.BLUE;
        return StdDraw.BLACK;
    }
}
Election Map
Election Map Data Type

**ElectionMap.** Represents the election map for a given election.

```java
public static void main(String[] args) {
    String name = args[0];
    int year = Integer.parseInt(args[1]);
    ElectionMap election = new ElectionMap(name, year);
    election.show();
}
```

% java ElectionMap NJ 2008
% java ElectionMap USA 1968
public class ElectionMap {
    private final int N;
    private final Region[] regions;
    private final VoteTally[] votes;

    public ElectionMap(String name, int year) {
        In in = new In(name + " .txt");

        // read in bounding box and rescale coordinates
        N = in.readInt();
        regions = new Region[N];
        votes = new VoteTally[N];
        for (int i = 0; i < N; i++) {
            String name = in.readLine();
            String usps = in.readLine();
            Polygon poly = new Polygon(in);
            regions[i] = new Region(name, usps, poly);
            votes[i] = new VoteTally(name, usps, year);
        }
    }

    public void show() {
        for (int i = 0; i < N; i++) {
            StdDraw.setPenColor(votes[i].getColor());
            regions[i].draw();
        }
    }
}
Modular Programming

Modular program. Collection of data types.
Data Visualization
Visual Display of Quantitative Information

Red states, blue states. Creates a misleading and polarizing picture.

Edward Tufte. Create charts with high data density that tell the truth.
**Purple America**

**Idea.** [Robert J. Vanderbei] Assign color based on number of votes.
- $a_1 =$ McCain votes.
- $a_2 =$ Other votes.
- $a_3 =$ Obama votes.

$$\begin{align*}
(R, G, B) &= \left(\frac{a_1}{a_1 + a_2 + a_3}, \frac{a_2}{a_1 + a_2 + a_3}, \frac{a_3}{a_1 + a_2 + a_3}\right)
\end{align*}$$

**Implementation.** Change one method!

```java
public Color getColor() {
    int tot = dem + rep + ind;
    return new Color((float) rep/tot, (float) ind/tot, (float) dem/tot);
}
```
Purple New Jersey

% java ElectionMap NJ 2004  % java ElectionMap NJ 2008
Purple America

% java ElectionMap USA 2008
Purple America

% java ElectionMap USA-county 2008
Data Visualization: Design Issues

**Remark.** Humans perceive red more strongly than blue.

**Remark.** Amount of color should be proportional to number of votes, not geographic boundary.

**Remark.** Project latitude + longitude coordinates to 2d plane.

Mercator projection

Albers projection
3D Visualization

3D visualization. Volume proportional to votes; azimuthal projection.
This map, which uses the size of circles to indicate a winning margin, shows how Senator Kerry had huge margins in many counties with big cities, while President Bush had relatively similar but consistent margins in suburban and rural counties.

Largest vote margin for Kerry: Cook County, Ill. (+805,857 votes)

Largest vote margin for Bush: Orange County, Calif. (+155,010 votes)

Second largest vote margin for Bush: Tarrant County, Tex. (+113,163 votes)
Cartograms

**Cartogram.** Area of state proportional to number of electoral votes.
Cartograms

*Cartogram.* Area of country proportional to population.
Summary

Modular programming.
- Break a large program into smaller independent components.
- Develop a data type for each component.
- Ex: Polygon, Region, VoteTally, ElectionMap, In, Out.

Ex 1. Build large software project.
- Software architect specifies API.
- Each programmer implements one module.
- Debug and test each piece independently. [unit testing]

Ex 2. Build reusable libraries.
- Language designer extends language with new data types.
- Programmers share extensive libraries.
- Ex: In, Out, Draw, Polygon, ...

Data visualization. You can do it! (worthwhile to learn from Tufte)