

# Computer-Assisted Reporting: Screen Scraping, Mapping in R

SPJ Region 7 Spring Conference  
Overland Park, KS  
April 9, 2011

Earl F. Glynn  
Overland Park, KS

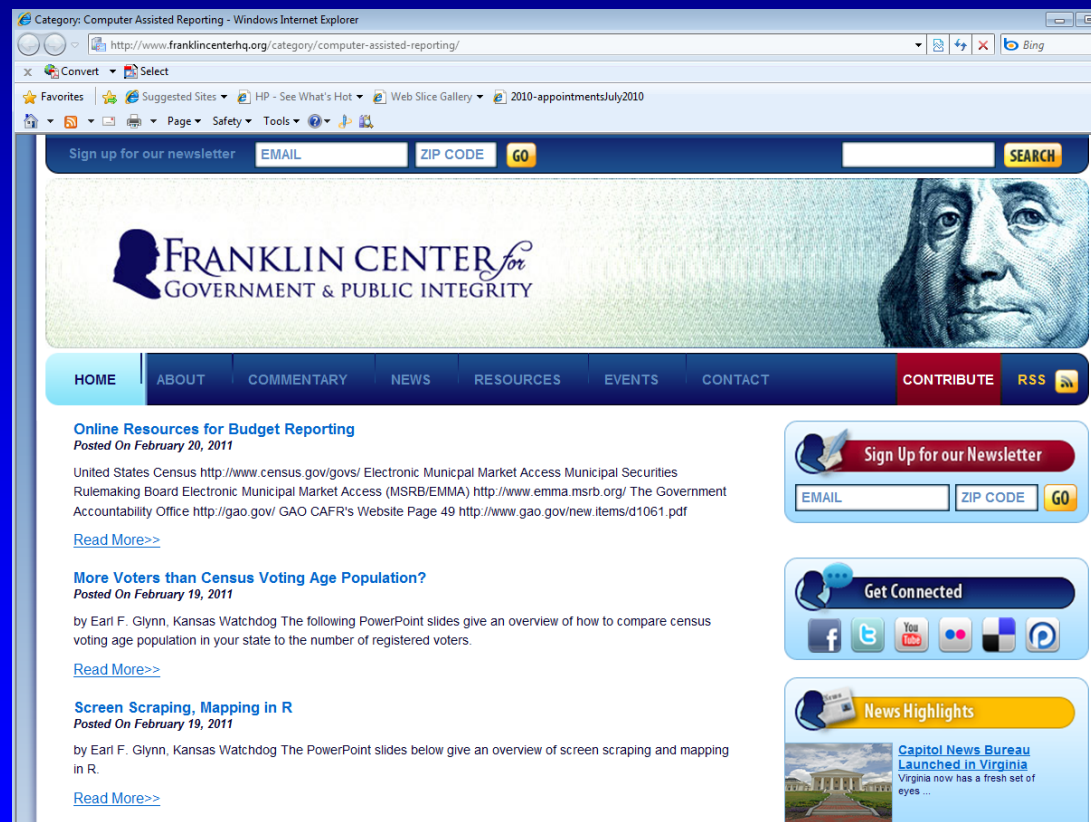
 FRANKLIN CENTER *for*  
GOVERNMENT & PUBLIC INTEGRITY



**KANSAS**  
**WATCHDOG**  .ORG™

# Computer-Assisted Reporting Blog

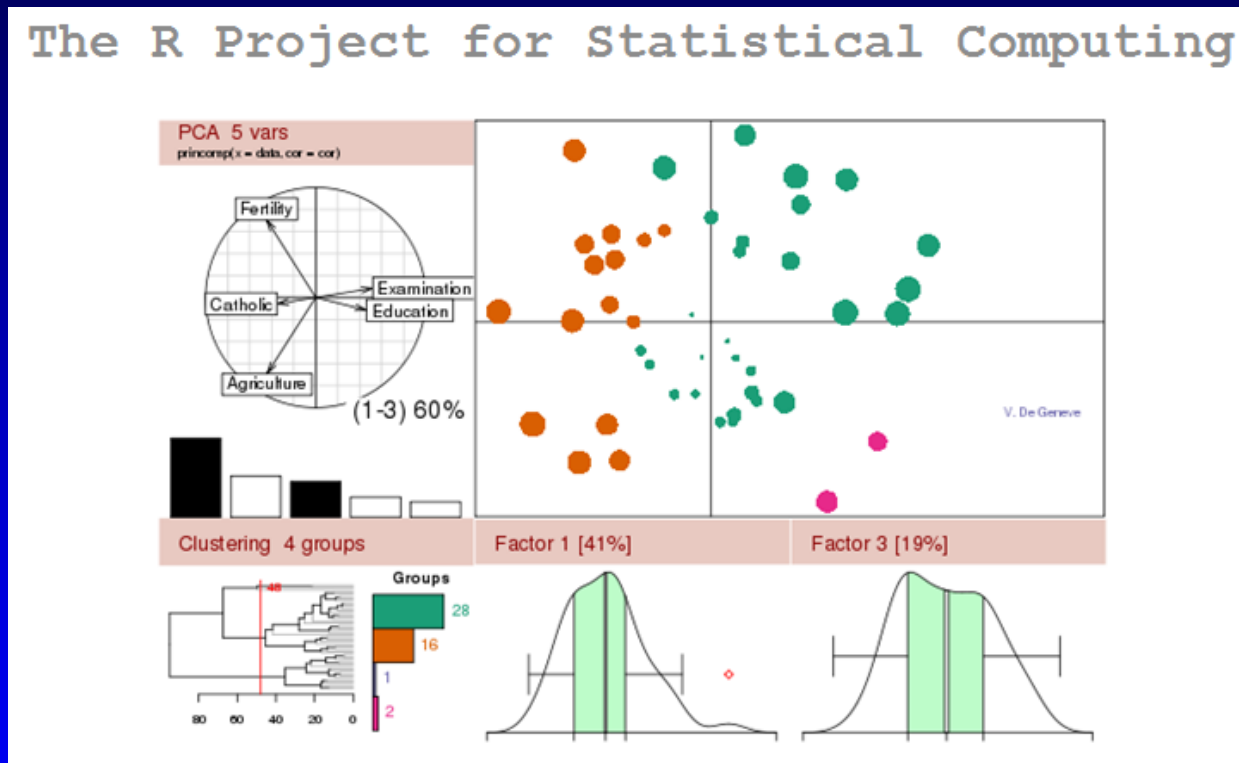
<http://car.franklincenterhq.org>



# Screen Scraping Example

- R Data Tool
- R Screen Scraping: 105 counties of data
- R Reshape Example
- Analysis/Mapping KS Judicial Retention Elections by County Using R
- Other Screen Scraping Examples

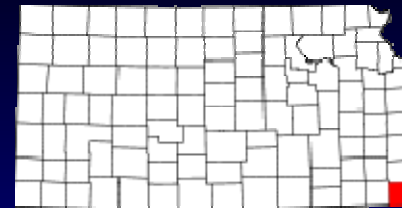
# Use to scrape web page



<http://www.r-project.org/>

Free tool for data manipulation, analysis,  
statistics and charts

# Screen Scraping




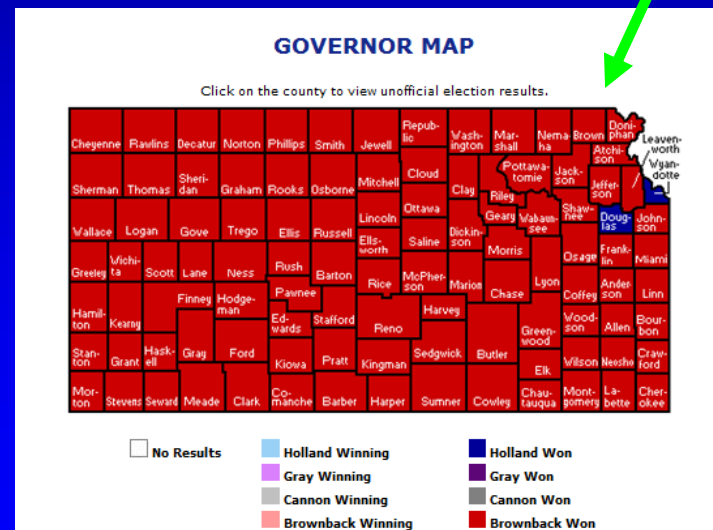
**Problem:** Visual election results by county not available for all statewide contests, and in particular, judicial retention elections.

**Solution:** "Screen scrape" 105 web pages (one per county), create table, plot map.

# Visual Election Results by County



Governor / Lt. Governor				
Precincts Reporting: 3315 of 3315				
Candidate	Votes	%		View Map
D-Tom Holland	264,214	32%	<div></div>	
L-Andrew P. Gray	21,932	3%	<div></div>	
F-Kenneth (Ken) W. Cannon	15,050	2%	<div></div>	
R-Sam Brownback	522,540	63%	<div></div>	



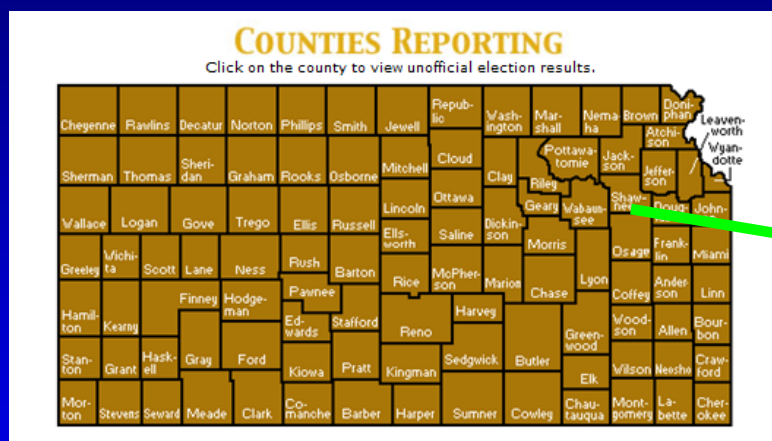
[www.kssos.org/ent/kssos\\_ent.html](http://www.kssos.org/ent/kssos_ent.html)

... but not for all statewide contests

Supreme Court Justice- 01				
Precincts Reporting:3315 of 3315				
Candidate	Votes	%		
Carol A. Beier - "YES"	439,474	63%	<div></div>	
Carol A. Beier - "NO"	256,806	37%	<div></div>	
<a href="#">Back to Top</a>				
Supreme Court Justice- 02				
Precincts Reporting:3315 of 3315				
Candidate	Votes	%		
Dan Biles - "YES"	424,952	62%	<div></div>	
Dan Biles - "NO"	261,169	38%	<div></div>	
<a href="#">Back to Top</a>				
Supreme Court Justice- 03				
Precincts Reporting:3315 of 3315				
Candidate	Votes	%		
Lawton R. Nuss - "YES"	428,828	63%	<div></div>	
Lawton R. Nuss - "NO"	257,019	38%	<div></div>	
<a href="#">Back to Top</a>				
Supreme Court Justice- 05				
Precincts Reporting:3315 of 3315				
Candidate	Votes	%		
Marla J. Luckert - "YES"	428,714	63%	<div></div>	
Marla J. Luckert - "NO"	255,057	37%	<div></div>	
<a href="#">Back to Top</a>				

Were Supreme Court justices retained in all counties?

# Data available online by county



[http://www.kssos.org/ent/kssos\\_ent.html](http://www.kssos.org/ent/kssos_ent.html)

Can click on each county  
to get results

Supreme Court Justice- 01				
County Precincts Reporting:0201 of 0201				
State Precincts Reporting:3315 of 3315				
Candidate	County Votes	County %	State Votes	State %
Carol A. Beier - "YES"	33,016	65%	439,474	63%
Carol A. Beier - "NO"	17,845	35%	256,806	37%
<a href="#">Back to Top</a>				

Supreme Court Justice- 02				
County Precincts Reporting:0201 of 0201				
State Precincts Reporting:3315 of 3315				
Candidate	County Votes	County %	State Votes	State %
Dan Biles - "YES"	31,507	63%	424,952	62%
Dan Biles - "NO"	18,646	37%	261,169	38%
<a href="#">Back to Top</a>				

Supreme Court Justice- 03				
County Precincts Reporting:0201 of 0201				
State Precincts Reporting:3315 of 3315				
Candidate	County Votes	County %	State Votes	State %
Lawton R. Nuss - "YES"	31,080	62%	428,828	63%
Lawton R. Nuss - "NO"	18,964	38%	257,019	38%
<a href="#">Back to Top</a>				

Supreme Court Justice- 05				
County Precincts Reporting:0201 of 0201				
State Precincts Reporting:3315 of 3315				
Candidate	County Votes	County %	State Votes	State %
Marla J. Luckert - "YES"	33,251	65%	428,714	63%
Marla J. Luckert - "NO"	17,571	35%	255,057	37%
<a href="#">Back to Top</a>				

[www.kssos.org/ent/shawnee.html](http://www.kssos.org/ent/shawnee.html)



## Visit 105 Web Pages to Analyze All Data

[www.kssos.org/ent/allen.html](http://www.kssos.org/ent/allen.html)

[www.kssos.org/ent/anderson.html](http://www.kssos.org/ent/anderson.html)

[www.kssos.org/ent/atchison.html](http://www.kssos.org/ent/atchison.html)

.

.

.

[www.kssos.org/ent/wilson.html](http://www.kssos.org/ent/wilson.html)

[www.kssos.org/ent/woodson.html](http://www.kssos.org/ent/woodson.html)

[www.kssos.org/ent/wyandotte.html](http://www.kssos.org/ent/wyandotte.html)

Need Automation

[www.kssos.org/ent/shawnee.html](http://www.kssos.org/ent/shawnee.html)

Supreme Court Justice- 01				
County Precincts Reporting:0201 of 0201				
State Precincts Reporting:3315 of 3315				
Candidate	County Votes	County %	State Votes	State %
Carol A. Beier - "YES"	33,016	65%	439,474	63%
Carol A. Beier - "NO"	17,845	35%	256,806	37%

# Web page

```
<tr><td align="center"><a href="#top">Back to Top</td></tr></table>  
  
<table width="500" border="0" cellspacing="0" cellpadding="0">  
<tr><td width="500" height="20" align="left" valign="top"><table width="500" border="0" cellpadding="0" cellspacing="2" bgcolor="#e1b85b">  
<tr><td width="500" align="center" valign="top"><A NAME="#018"></A><A HREF="#018"></A><span class="headers">Supreme Court Justice- 01</span><br />  
<span class="result_numbers">County Precincts Reporting:020 of 0201</span><br />  
<span class="result_numbers">State Precincts Reporting:3315 of 3315</span></td></tr></table></td></tr>  
<tr><td width="500" align="left" valign="top"><table width="500" border="0" cellpadding="0" cellspacing="2" bgcolor="#FFE37E">  
<tr><td width="181" height="20" align="left" valign="middle" class="racetitle">Candidate</td>  
<td width="79" height="20" align="right" valign="middle" class="racetitle">County<br />Votes</td>  
<td width="55" height="20" align="center" valign="middle" class="racetitle">County<br />%</td>  
<td width="79" height="20" align="right" valign="middle" class="racetitle">State<br />Votes</td>  
<td width="55" height="20" align="center" valign="middle" class="racetitle">State<br />%</td>  
<td width="95" height="20" align="center" valign="middle" class="racetitle">  
nbsp:</td></tr></table></td></tr>  
<tr><td width="500" align="left" valign="top"><table width="500" border="0" cellpadding="0" cellspacing="2" bgcolor="#e3e3e3">  
<tr><td width="181" align="left" valign="middle" class="result_numbers"><strong>nbspnbspn Carol A. Beier - "YES"</strong></td>  
<td width="70" align="right" valign="middle" class="result_numbers">33,016</td>  
<td width="55" align="center" valign="middle" class="result_numbers">65%</td>  
<td width="70" align="right" valign="middle" class="result_numbers">439,474</td>  
<td width="55" align="center" valign="middle" class="result_numbers">63%</td>  
<td width="69" align="center" valign="middle" class="result_numbers">nbspn:</td>  
</tr></table></td></tr>  
<tr><td width="500" align="left" valign="top"><table width="500" border="0" cellpadding="0" cellspacing="2" bgcolor="#e3e3e3">  
<tr><td width="181" align="left" valign="middle" class="result_numbers"><strong>nbspnbspn Carol A. Beier - "NO"</strong></td>  
<td width="70" align="right" valign="middle" class="result_numbers">17,845</td>  
<td width="55" align="center" valign="middle" class="result_numbers">35%</td>  
<td width="70" align="right" valign="middle" class="result_numbers">256,806</td>  
<td width="55" align="center" valign="middle" class="result_numbers">37%</td>  
<td width="69" align="center" valign="middle" class="result_numbers">nbspn:</td>  
</tr></table></td></tr>
```

HTML

# HTML

IE: Right Click, View Source  
Firefox: Right Click, View Page Source

# Screen scrape: Web page to data file

Supreme Court Justice- 01				
County Precincts Reporting:0201 of 0201				
State Precincts Reporting:3315 of 3315				
Candidate	County Votes	County %	State Votes	State %
Carol A. Beier - "YES"	33,016	65%	439,474	63%
Carol A. Beier - "NO"	17,845	35%	256,806	37%



Parse HTML

```
Shawnee|Supreme Court Justice- 01|Candidate|CountyVotes|County|StateVotes|State
Shawnee|Supreme Court Justice- 01|Carol A. Beier - YES|33016|65|439474|63
Shawnee|Supreme Court Justice- 01|Carol A. Beier - NO|17845|35|256806|37
```

+ Data Cleanup

+ Filtering

# "Reshape" Data

Reshape from "long" to "wide" format

County	Contest	Candidate	CountyVotes	CountyPercent
Allen Supreme Court Justice- 01	Carol A. Beier - YES	2396	63	
Allen Supreme Court Justice- 01	Carol A. Beier - NO	1436	38	
Anderson Supreme Court Justice- 01	Carol A. Beier - YES	1264	54	
Anderson Supreme Court Justice- 01	Carol A. Beier - NO	1091	46	
Atchison Supreme Court Justice- 01	Carol A. Beier - YES	2272	58	
Atchison Supreme Court Justice- 01	Carol A. Beier - NO	1655	42	

```
> wide <- reshape(court1, v.names=c("CountyVotes", "CountyPercent"),
+                   timevar="Candidate",
+                   idvar="County", direction="wide")
```

County	Contest	Votes.Carol A. Beier - YES	Pct.Carol A. Beier - YES	Votes.Carol A. Beier - NO	Pct.Carol A. Beier - NO
Allen Supreme Court Justice- 01		2396	63	1436	38
Anderson Supreme Court Justice- 01		1264	54	1091	46
Atchison Supreme Court Justice- 01		2272	58	1655	42
Barber Supreme Court Justice- 01		966	56	769	44
Barton Supreme Court Justice- 01		4505	62	2732	38
Bourbon Supreme Court Justice- 01		2429	58	1736	42

# Analyze Data

## "Worst" Counties (< 50%)

```
> Carol.Beier[Carol.Beier$Beier.Yes <= Carol.Beier$Beier.No, ]
```

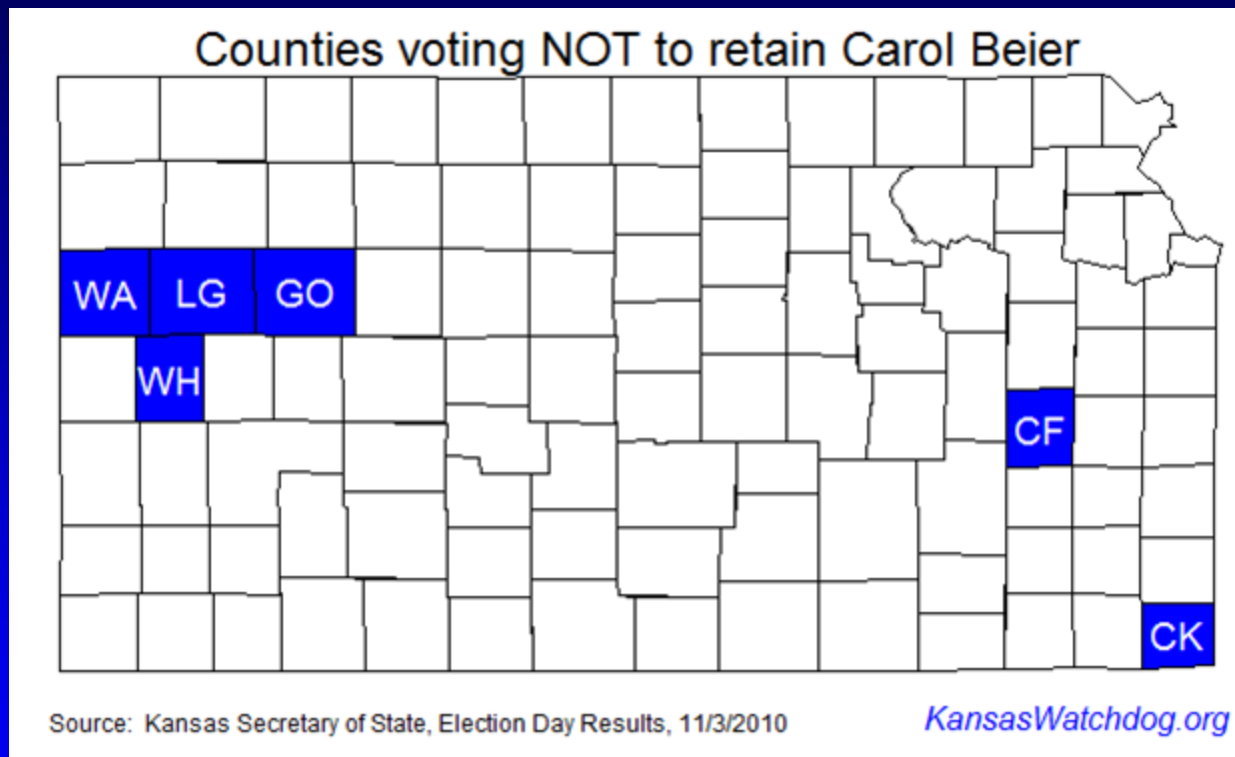
	County	Beier.Yes	Beier.No	Beier.PercentYes
11	Cherokee	2414	2784	46.4
16	Coffey	1082	1248	46.4
32	Gove	429	433	49.8
55	Logan	397	486	45.0
100	Wallace	200	296	40.3
102	Wichita	289	326	47.0

## "Best" Counties (Top 5)

```
> Carol.Beier[order(Carol.Beier$Beier.PercentYes, decreasing=TRUE)[1:5], ]
```

	County	Beier.Yes	Beier.No	Beier.PercentYes
23	Douglas	19802	7104	73.6
81	Riley	8506	3551	70.5
105	Wyandotte	17293	7578	69.5
31	Geary	2934	1306	69.2
85	Saline	10730	4807	69.1

# Map Data



# Article



## 10 counties voted not to retain Kansas Supreme Court justices

By Earl Glynn On November 10, 2010

Unofficial election night results from the Kansas Secretary of State last week showed all four Kansas Supreme Court justices on the ballot were retained with about 62% voting "yes" and about 38% voting "no" to retain the judges. But not all counties voted to retain.

The Supreme Court judges facing retention elections this year were [Carol Beier](#), [Dan Biles](#), [Lawton Nuss](#) and [Marla Luckert](#).

Kansas Watchdog analyzed the results by county, which showed 10 counties voted to not retain one or more of the judges out of the 105 counties in the state.

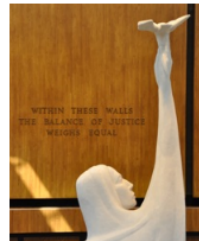
Five counties (Cherokee, Coffey, Logan, Wallace and Wichita) voted "no" to all four judges.

### Kansas counties voting not to retain Supreme Court Justices

County	Abbr	Kansas Supreme Court Justice in 2010 Retention Election			
		Beier	Biles	Luckert	Nuss
Anderson	AN		Tie		
Cherokee	CK	X	X	X	X
Cheyenne	CN		X	X	
Coffey	CF	X	X	X	X
Doniphan	DP		X		X
Franklin	FR		X		
Gove	GO	X		X	
Logan	LG	X	X	X	X
Rawlins	RA			X	
Wallace	WA	X	X	X	X
Wichita	WH	X	X	X	X

"X" = County voted NOT to retain

[Print This Article](#)



Statue of Justice at Kansas Judicial Center



# Resource



<http://www.propublica.org/nerds/item/doc-dollars-guides-collecting-the-data>



# Questions?

## Earl F. Glynn

[earl@KansasWatchdog.org](mailto:earl@KansasWatchdog.org)

Twitter: @KansasWatchdog



[earl.glynn@FranklinCenterHq.org](mailto:earl.glynn@FranklinCenterHq.org)

