

Processing the Florida Early Voting Files:

Paul Gronke

Reed College

Early Voting Information Center

Florida Early Voting Files

- Files are stored on a daily basis but in a complex format
 - 67 separate files
 - No pattern to the names
 - Individual file names are not an easily identified URL
- Downloading is a challenge: need to use a web “harvester” or “scraper.”
- Where to start:
 - Florida Division of Elections URL
 - <http://election.dos.state.fl.us/>
 - Florida Early Voting Report URL
 - <https://countyballotreports.elections.myflorida.com/AbsenteeEarlyVotingReports/PublicReports>

A Field Guide to the Website Harvest

- Files are listed by Election and grouped by county

- 67 x 3 separate files on this page

- File and variable documentation are at the top of the page

- Right click on an individual file to recover the URL

The screenshot shows the Florida Division of Elections website. At the top left is the Great Seal of the State of Florida. The main header reads 'Absentee Ballot Request and Early Voting Reports' and 'Florida Division of Elections'. A navigation bar includes 'Home', 'Voting Information Index', 'Voter Registration Index', and 'Administration'. The main content area is titled 'Early Voting Reports' and contains a list of links to various documents: 'Table of Election Numbers(pdf, 11KB)', 'Table of County Codes(pdf, 19KB)', 'Table of Party Codes(pdf, 13KB)', 'Table of Absentee Request Status Codes(pdf, 10KB)', and 'File Layout Specification for the 3 file types presented(pdf, 18KB)'. A blue box labeled 'File documentation' has an arrow pointing to the 'Table of Party Codes' link. Below this is a disclaimer. A blue box titled 'The 2012 General Election file for Alachua County' contains a table of election data:

Election ID	Election Description	Download archive
10161	Election 10161	Download archive
9546	2012 Primary Election	Download archive
9547	2012 General Election	Download archive
9690	2012 Presidential Preference Primary	Download archive

Below the table is a section for 'County Early Voting Reports' with tabs for 'Early Voting Lists' and 'Early Voting Summaries'. The 'Early Voting Summaries' tab is active, showing a table with columns for 'County', 'Election ID', 'Election', 'Last Updated', and 'File Size':

County	Election ID	Election	Last Updated	File Size
ALACHUA	9546	2012 Primary Election	8/13/2012 12:47:04 AM	596 KB
ALACHUA	9547	2012 General Election	10/30/2012 12:48:00 AM	1,557 KB

Identifying your prey

- The URL for the Alachua file is
 - https://countyballotreports.elections.myflorida.com/AbsenteeEarlyVotingReports/ViewReport?name=https%3A%2F%2Ffdos1fvrs.blob.core.windows.net%2Fcountyballotreport%2FALA_EVL_9547_10302012_004800000
- Wow!
 - The final section of the URL starting with “ALA EVL” identifies this as the Alachua file
 - The string that reads “10302012” identifies this as the November 2012 election file
 - No regular pattern to the 10 characters at the end. This will make “batch” downloading a challenge
- You have 67 separate files to download!

Download options

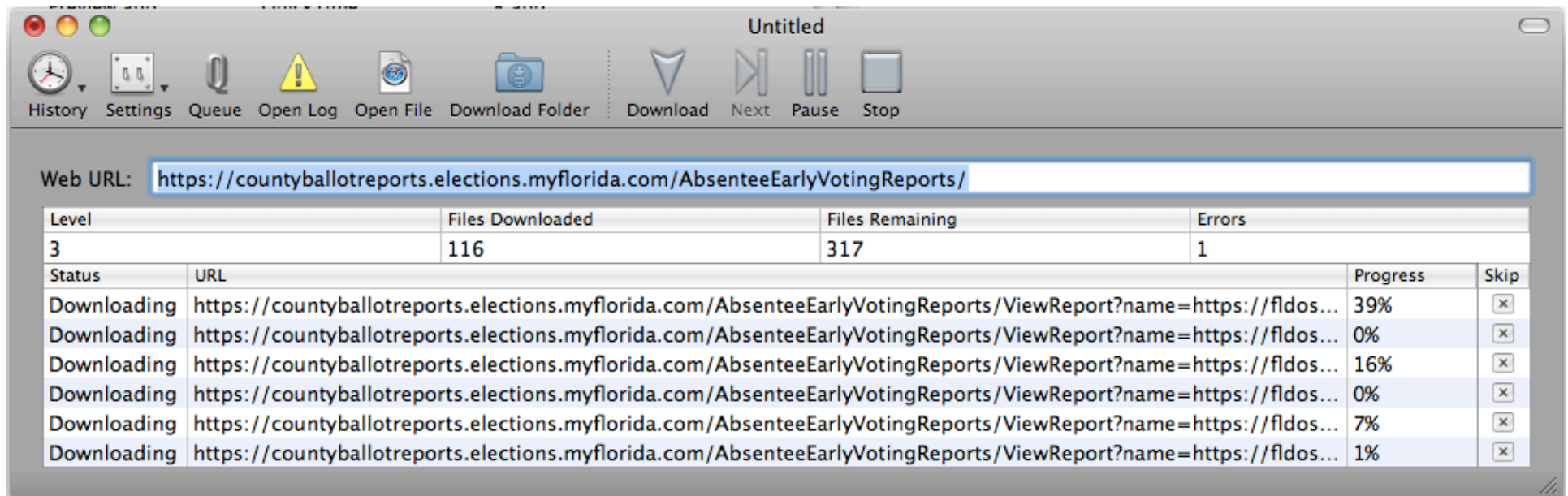
Downloading 67 large files on a regular basis

1. Download them manually. This is a job for an intern!
2. My advice: download them using a “batch” downloading program
 - Mac users: use SiteSucker (free software)
<http://sitesucker.us/home.html>
 - Windows users: Teleport Pro is a low-priced option (there may be other free options). <http://www.tenmax.com/teleport/pro/home.htm>
3. Experts may be able to write download scripts or from other utilities*

* Wget (Linux) or cURL (Mac OSX) can both recover files using the Florida URL but can't use wildcards. A 67 line script is not difficult to write but you need the list of three letter county codes as well as the nine-digit string at the end of the URL.

Sitesucker Example

- Using the \root early/absentee URL will download all the relevant files
 - <https://countyballotreports.elections.myflorida.com/AbsenteeEarlyVotingReports/PublicReports>)
- Sitesucker in operation on the Florida files:



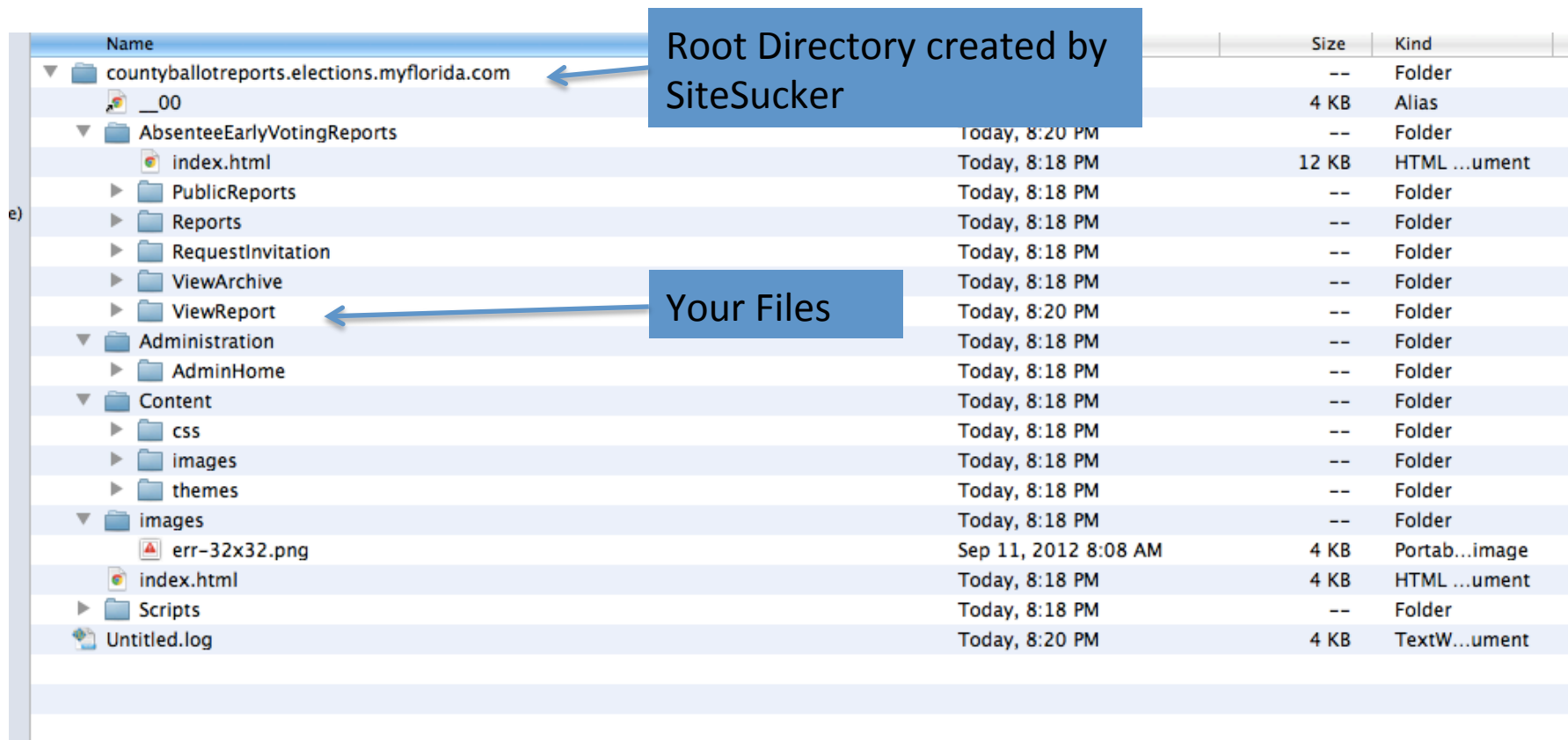
The screenshot shows the Sitesucker application window titled "Untitled". The interface includes a toolbar with icons for History, Settings, Queue, Open Log, Open File, Download Folder, Download, Next, Pause, and Stop. The "Web URL:" field contains the text: <https://countyballotreports.elections.myflorida.com/AbsenteeEarlyVotingReports/>. Below the URL field is a table displaying the download progress.

Level	Files Downloaded	Files Remaining	Errors
3	116	317	1

Status	URL	Progress	Skip
Downloading	https://countyballotreports.elections.myflorida.com/AbsenteeEarlyVotingReports/ViewReport?name=https://fdos...	39%	<input type="checkbox"/>
Downloading	https://countyballotreports.elections.myflorida.com/AbsenteeEarlyVotingReports/ViewReport?name=https://fdos...	0%	<input type="checkbox"/>
Downloading	https://countyballotreports.elections.myflorida.com/AbsenteeEarlyVotingReports/ViewReport?name=https://fdos...	16%	<input type="checkbox"/>
Downloading	https://countyballotreports.elections.myflorida.com/AbsenteeEarlyVotingReports/ViewReport?name=https://fdos...	0%	<input type="checkbox"/>
Downloading	https://countyballotreports.elections.myflorida.com/AbsenteeEarlyVotingReports/ViewReport?name=https://fdos...	7%	<input type="checkbox"/>
Downloading	https://countyballotreports.elections.myflorida.com/AbsenteeEarlyVotingReports/ViewReport?name=https://fdos...	1%	<input type="checkbox"/>

Sitesucker Example (2)

- The files will be stored in an unusual location. The other directories and files are not needed



Name	Size	Kind
countyballotreports.elections.myflorida.com	--	Folder
_00	4 KB	Alias
AbsenteeEarlyVotingReports	Today, 8:20 PM	Folder
index.html	Today, 8:18 PM	12 KB HTML ...ument
PublicReports	Today, 8:18 PM	-- Folder
Reports	Today, 8:18 PM	-- Folder
RequestInvitation	Today, 8:18 PM	-- Folder
ViewArchive	Today, 8:18 PM	-- Folder
ViewReport	Today, 8:20 PM	-- Folder
Administration	Today, 8:18 PM	-- Folder
AdminHome	Today, 8:18 PM	-- Folder
Content	Today, 8:18 PM	-- Folder
css	Today, 8:18 PM	-- Folder
images	Today, 8:18 PM	-- Folder
themes	Today, 8:18 PM	-- Folder
images	Today, 8:18 PM	-- Folder
err-32x32.png	Sep 11, 2012 8:08 AM	4 KB Portab...image
index.html	Today, 8:18 PM	4 KB HTML ...ument
Scripts	Today, 8:18 PM	-- Folder
Untitled.log	Today, 8:20 PM	4 KB TextW...ument

SiteSucker (2)

- Files are given an HTML extension that needs to be replaced with TXT

Name	Date Modified	Size	Kind
countyballotreports.elections.myflorida.com	Today, 8:20 PM	--	Folder
__00	Today, 8:18 PM	4 KB	Alias
AbsenteeEarlyVotingReports	Today, 8:20 PM	--	Folder
index.html	Today, 8:18 PM	12 KB	HTML ...ument
PublicReports	Today, 8:18 PM	--	Folder
Reports	Today, 8:18 PM	--	Folder
RequestInvitation	Today, 8:18 PM	--	Folder
ViewArchive	Today, 8:18 PM	--	Folder
ViewReport	Today, 8:20 PM	--	Folder
index-name=https__fdos1fvrs.blob...VL_9546_08132012_004704000.html	Today, 8:18 PM	614 KB	HTML ...ument
index-name=https__fdos1fvrs.blob...VL_9547_10302012_004800000.html	Today, 8:18 PM	1.6 MB	HTML ...ument
index-name=https__fdos1fvrs.blob...VL_9690_01302012_000045000.html	Today, 8:18 PM	500 KB	HTML ...ument
index-name=https__fdos1fvrs.blob...VS_9546_08132012_004704000.html	Today, 8:20 PM	4 KB	HTML ...ument
index-name=https__fdos1fvrs.blob...VS_9547_10302012_004800000.html	Today, 8:20 PM	4 KB	HTML ...ument
index-name=https__fdos1fvrs.blob...VS_9690_01302012_000045000.html	Today, 8:20 PM	4 KB	HTML ...ument
index-name=https__fdos1fvrs.blob...VL_9546_08132012_004535000.html	Today, 8:18 PM	176 KB	HTML ...ument
index-name=https__fdos1fvrs.blob...VL_9547_10302012_005352000.html	Today, 8:18 PM	209 KB	HTML ...ument
index-name=https__fdos1fvrs.blob...VL_9690_01302012_000853000.html	Today, 8:18 PM	111 KB	HTML ...ument
index-name=https__fdos1fvrs.blob...VS_9546_08132012_004535000.html	Today, 8:20 PM	4 KB	HTML ...ument
index-name=https__fdos1fvrs.blob...VS_9547_10302012_005352000.html	Today, 8:20 PM	4 KB	HTML ...ument
index-name=https__fdos1fvrs.blob...VS_9690_01302012_000853000.html	Today, 8:20 PM	4 KB	HTML ...ument

Should have "txt" extension

Unix Gymnastics (Mac and Linux)

- Step 1: rename all files with an “.html” extension to “.txt”

```
paulgronke$ for i in *.html; do mv "$i" "${i/.html}"; done
```

- Step 2: Make a subdirectory below the the “ViewReport” directory
- Step 3: Copy only the November 2012 files to the subdirectory. All 11/2012 files have a string identifying the date the file was created (in this cases, “10202012”).

We can use this to copy the files to the subdirectory (this exact command only works if you are in a subdirectory below the directory where the early voting files are stored)

```
paulgronke$ find .. -name \*10302012\* -exec cp {} . \;
```

- NOTE: The string will change each day!
- How I learned this: <http://www.debian-administration.org/articles/150>

Finally ready for Stata

- The Stata Code linked here:
 - Issues a shell (Unix) command to creates a text file with the names of all individual Florida data file
 - Reads this file line by line as input into insheet commands
 - Appends the files one by one
 - Does some variable conversion and recoding
 - Collapses by date, summing up total Democratic, Republican, and Unaffiliated votes
 - Creates a simple bar chart

Commentary

- Florida does an excellent job making its early voting returns easily and cheaply available
- Process is still very complex and beyond the skills of experienced users
- If North Carolina is the best, and Florida is the second best, why is it so much harder (and expensive) in other states?