django

Database-driven journalism

Lawrence Journal-World

Lawrence, KS

“

Oh. The Django guys, right?

”

But we do so much more

Lifecycle of a “data story”

- ➜ Reporter gets data
- ➜ Parse/normalize data into RDBMS
- ➜ Build web interface
- ➜ Offer cleaned data for download

Python every step of the way

Real-life example

Crime reports on local university campus

Reporter got the data

- ➜ Three years of reports
- ➜ Organized by residence hall
- ➜ Tables embedded in Microsoft Word :(

Google to the rescue!

Google™
Docs
BETA

Step 1

- ➜ Upload MS Word to Google Docs
- ➜ Export as HTML

Step 2

- Google Docs produces almost-valid HTML
- BeautifulSoup gets us the rest of the way
- End result: a nicely-formatted CSV file

Data modelling

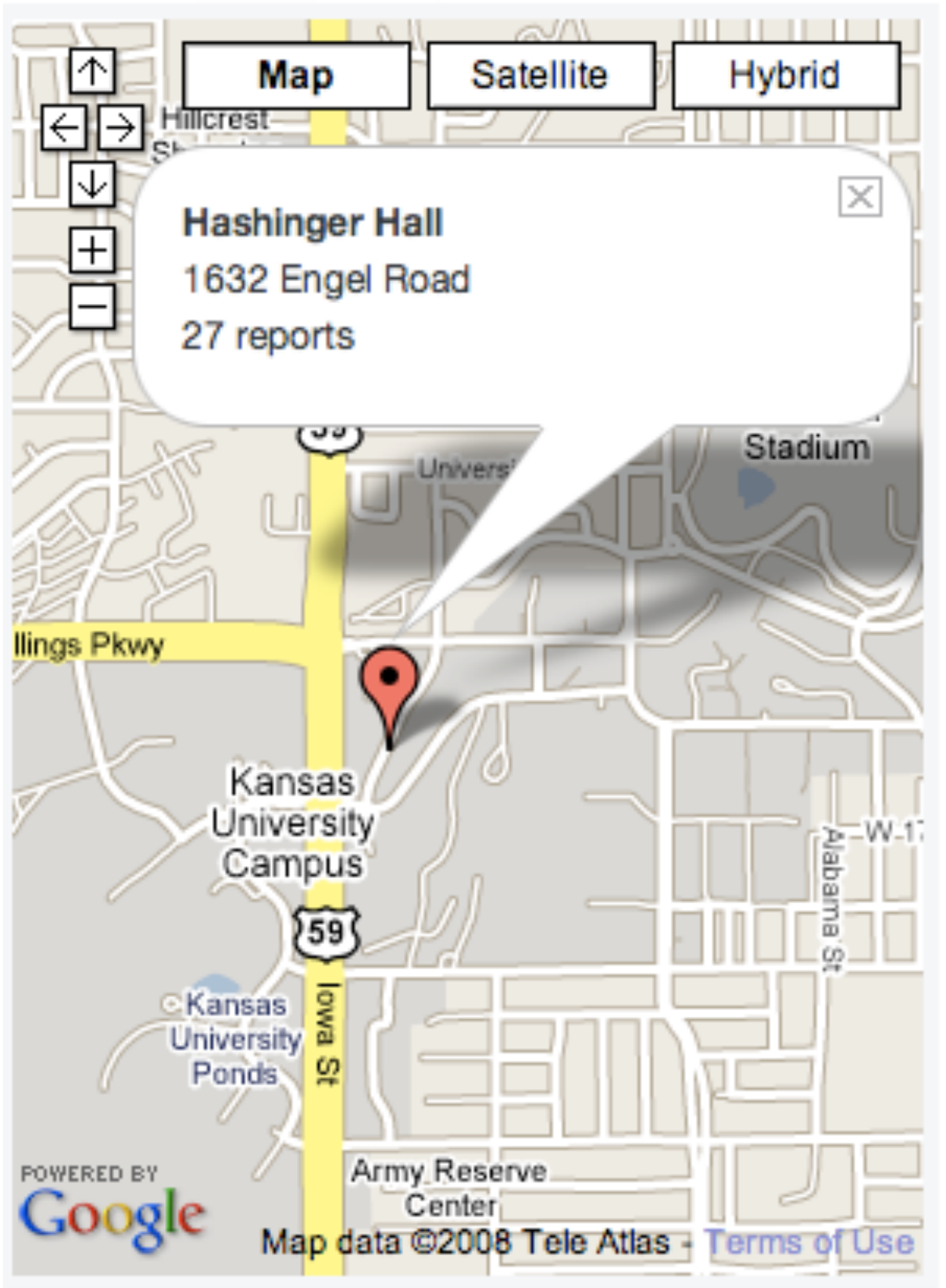
- ➜ Build a few classes: residence hall, crime, statute, etc.
- ➜ Import from the CSV files
- ➜ Used Django for this, but don't have to; any ORM is good, or just DB-API

Web-based views

- Pick your favorite web framework/toolkit
- We use Django
- Mostly you're slinging around lists of dictionaries

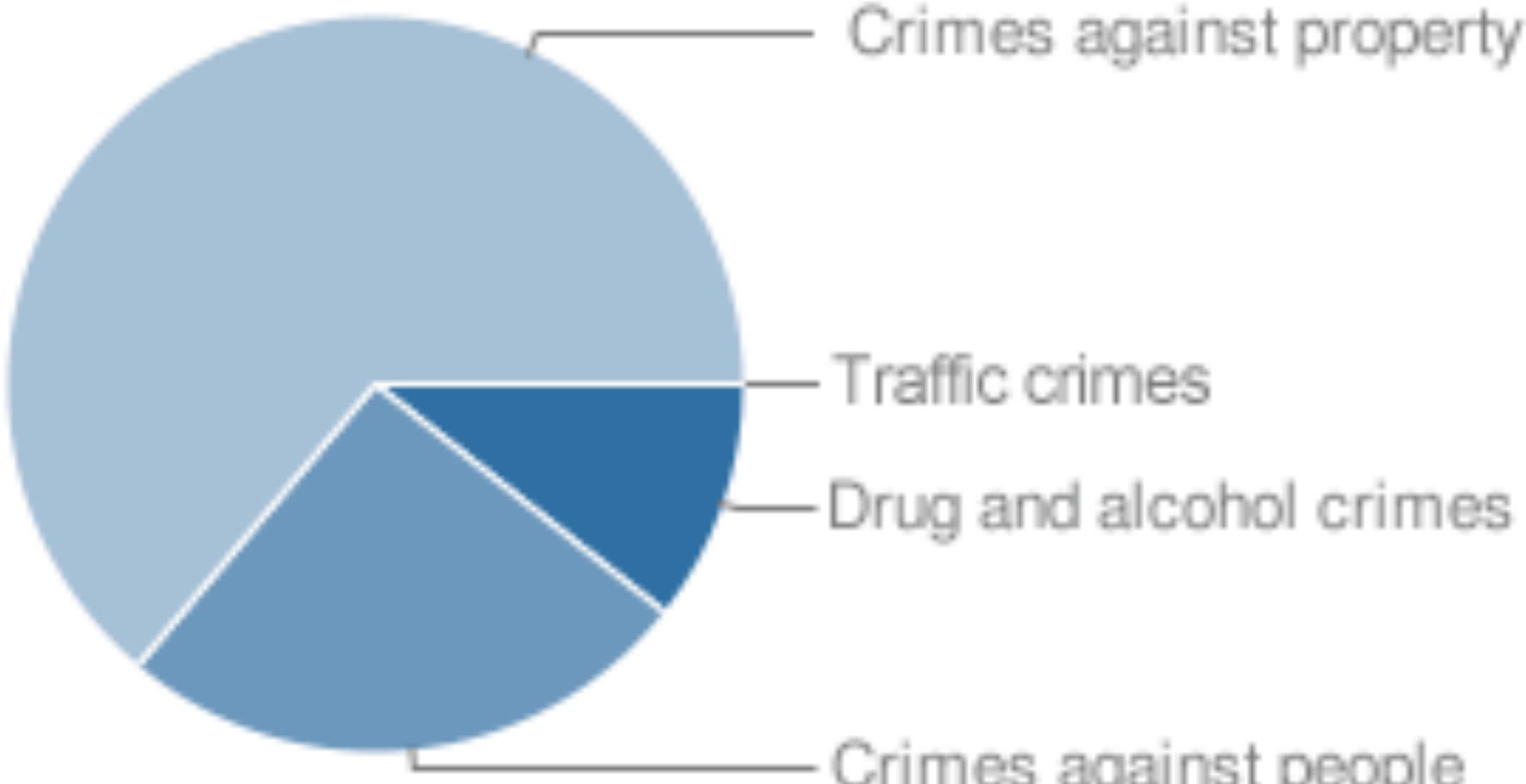
Make it fancy

- geocoding of addresses with geopy (http://exogen.case.edu/projects/geopy/)
- Now we have maps


Map
Satellite
Hybrid
Hillcrest
Hashinger Hall
1632 Engel Road
27 reports
Stadium
Universi
llings Pkwy
Kansas
University
Campus
59
Iowa St
Alabama St
W-17
Kansas
University
Ponds
Army Reserve
Center
POWERED BY
Google
Map data ©2008 Tele Atlas - Terms of Use

Make it fancy

- ➜ Google Charts API is just HTTP, we can talk to that with Python
- ➜ Now we have charts


Crimes against property
Traffic crimes
Drug and alcohol crimes
Crimes against people

And put it all together

Crimes reported in KU residences 2005–2007

Between **2005** and **2007**, **552** crimes were reported in KU residences, and a number of other crimes were reported on campus. Below, you can browse through the residence crime reports in a variety of ways: by **year**, by **residence hall**, and by **type of crime**. You can also **download the data** and examine it yourself, and **read the related story**.

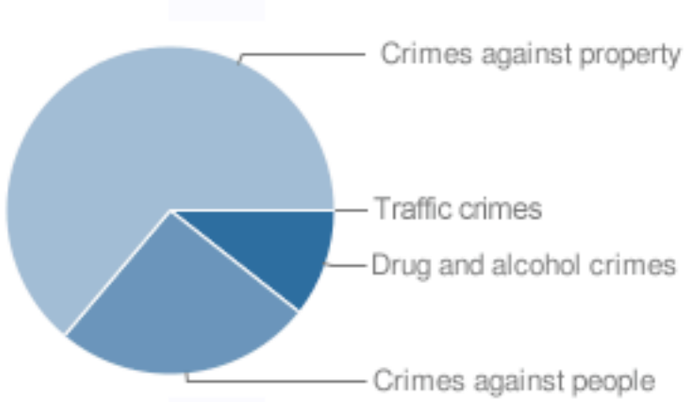
Browse reports by year:

2005	**198 reports**
2006	**185 reports**
2007	**169 reports**

Year	Reports
2005	198
2006	185
2007	169

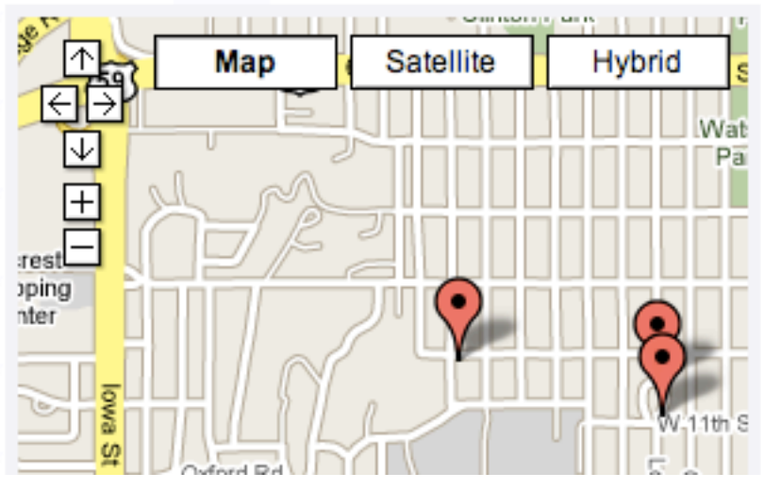
Browse reports by type:

Drug and alcohol crimes	**60 reports**
Crimes against people	**142 reports**
Crimes against property	**345 reports**
Traffic crimes	**5 reports**



Browse reports by residence:

Battenfeld Scholarship Hall	**3 reports**
Corbin Hall	**21 reports**
Dennis E. Rieger Scholarship Hall	**2 reports**
Douthart Scholarship Hall	**0 reports**
Ellsworth Hall	**45 reports**
Gertrude Sellards Pearson Hall	**17 reports**
Grace Pearson Scholarship Hall	**1 report**



http://
www2.ljworld.com/
data/crime/ku/

This took two days

- ➜ Got the data on Wednesday, showed off the interface on Friday
- ➜ Took less time to prepare this interface than to write the story for the newspaper
- ➜ Hooray Python!

We do this all the time

- Home foreclosures: http://www2.ljworld.com/foreclosures/sheriffs_sale/
- Use of university planes: http://www2.ljworld.com/transportation/air/
- Alcohol violations: http://www2.ljworld.com/data/alcohol_violations/mip/
- And more: http://www2.ljworld.com/special_coverage/

Python all the way

Get these slides:

http://www.b-list.org/

(and if you like this sort of stuff, we're hiring!)