PyData
Dallas 2015

There’s no place like home.

Analyzing HPD police beat data with Python and ArcGIS

Paige Bailey

@DynamicWebPaige
www.paige-bailey.com

Founder of PyLadies-HTX
GIS / Python at Chevron

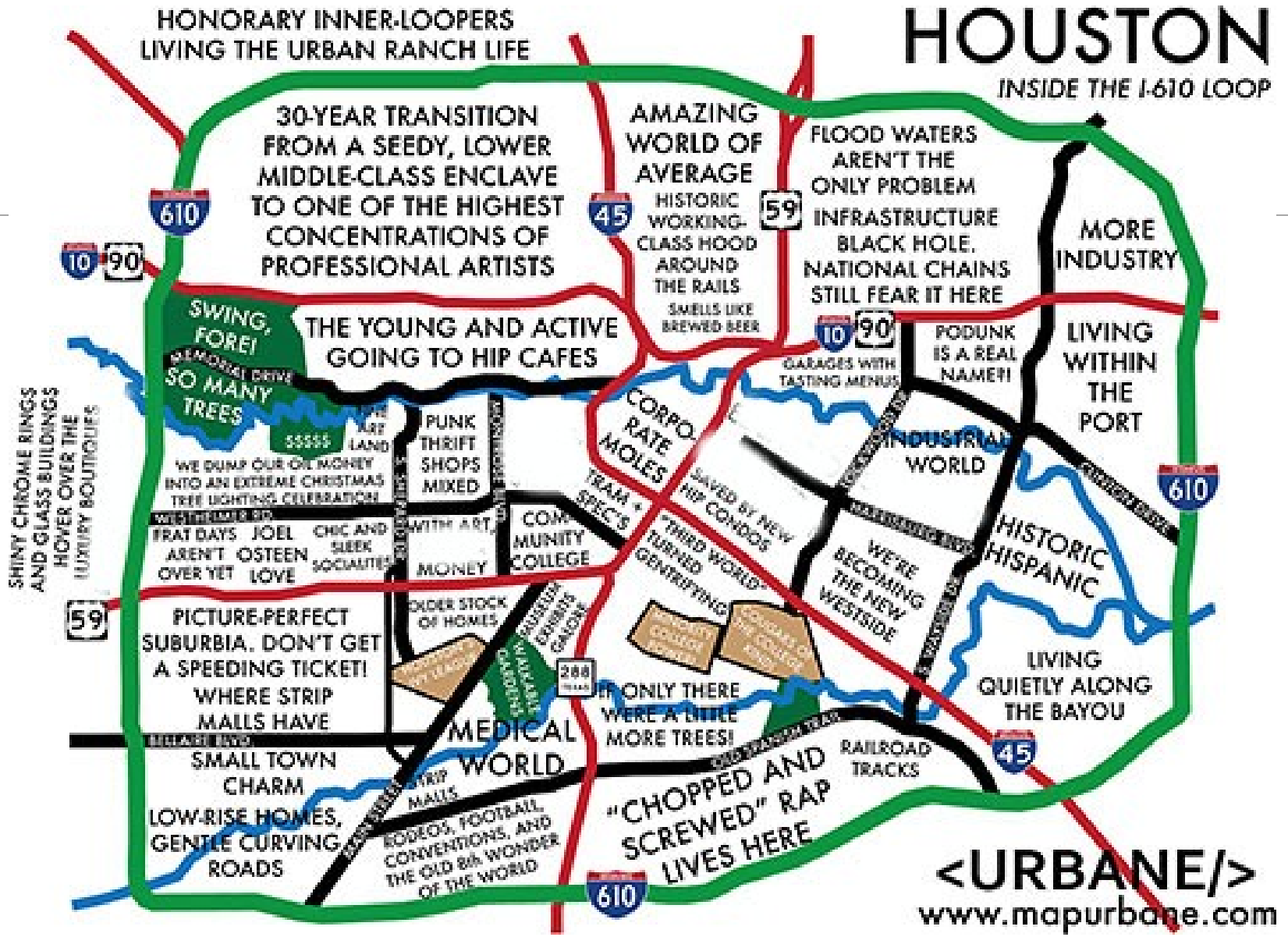
Houston, TX

B.S. Geophysics / B.A. Sociology
Rice University '13

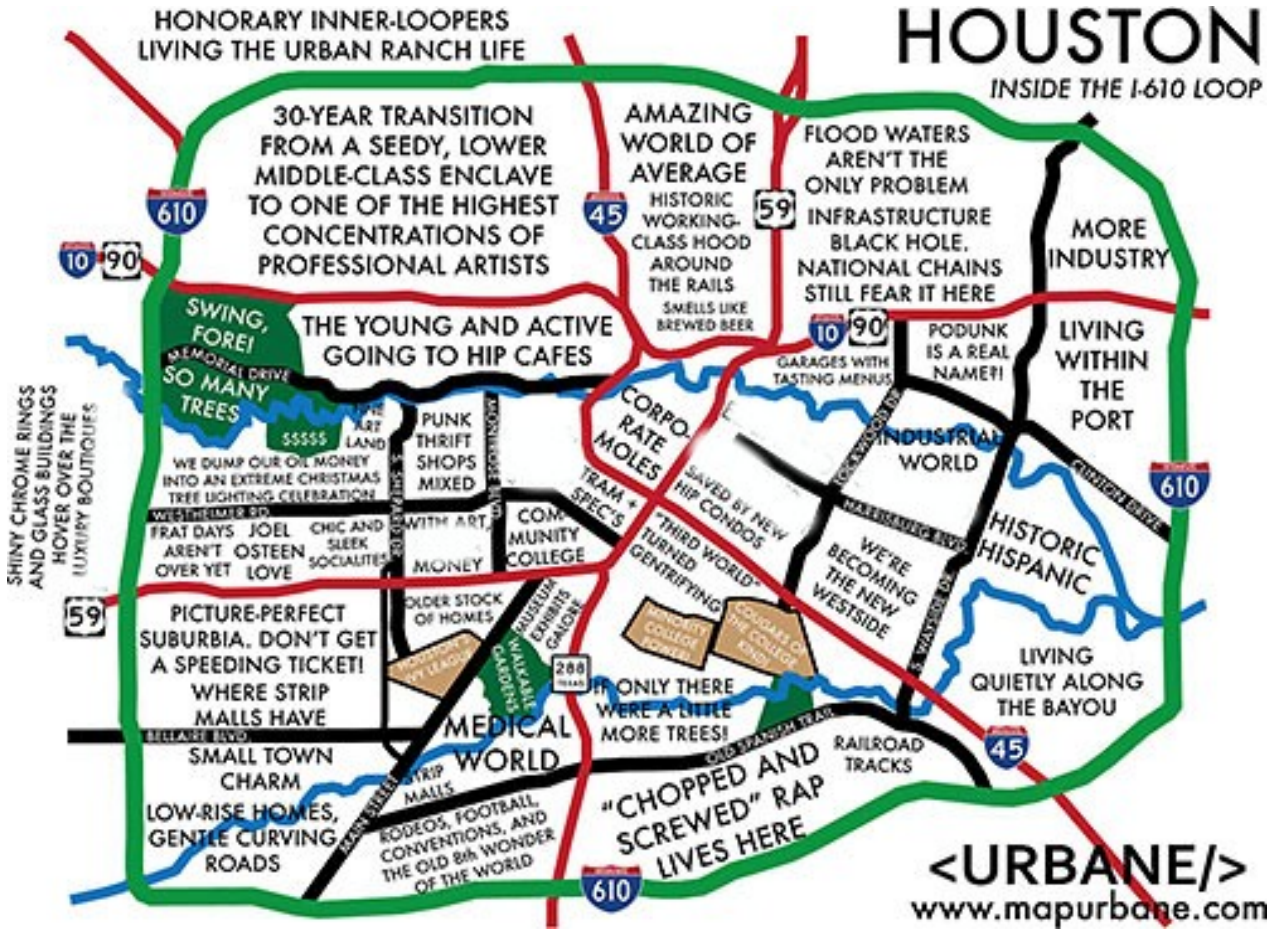
M.S. Subsurface Geoscience
Rice University '16

A.B.A Business Administration
Hill College '08

...apartment-hunting is a Thing.

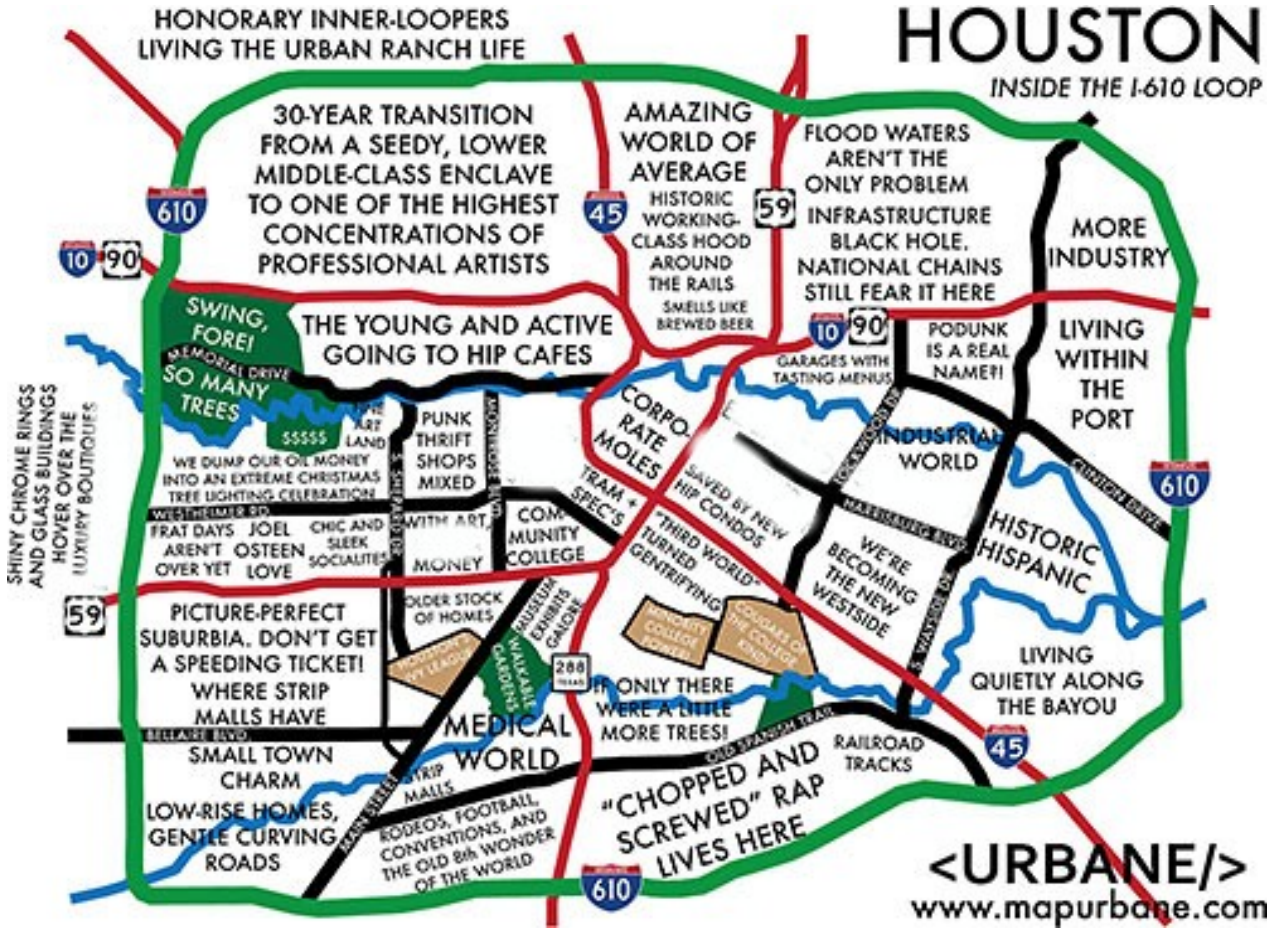

HONORARY INNER-LOOPERS
LIVING THE URBAN RANCH LIFE
HOUSTON
INSIDE THE I-610 LOOP
30-YEAR TRANSITION FROM A SEEDY, LOWER MIDDLE-CLASS ENCLAVE TO ONE OF THE HIGHEST CONCENTRATIONS OF PROFESSIONAL ARTISTS
AMAZING WORLD OF AVERAGE
HISTORIC WORKING-CLASS HOOD AROUND THE RAILS
SMELLS LIKE BREWED BEER
FLOOD WATERS AREN'T THE ONLY PROBLEM
INFRASTRUCTURE BLACK HOLE. NATIONAL CHAINS STILL FEAR IT HERE
MORE INDUSTRY
SWING, FORE!
SO MANY TREES
THE YOUNG AND ACTIVE GOING TO HIP CAFES
$$$$$
PUNK THRIFT SHOPS MIXED
WE DUMP OUR OIL MONEY INTO AN EXTREME CHRISTMAS TREE LIGHTING CELEBRATION
FRAT DAYS AREN'T OVER YET
JOEL OSTEEN LOVE
CHIC AND SLEEK SOCIALITES
MONEY
COMMUNITY COLLEGE
CORPORATE MOLES
SAVED BY NEW HIP CONDOS
"THIRD WORLD" TURNED GENTRIFYING
GARAGES WITH TASTING MENUS
PODUNK IS A REAL NAME?!
LIVING WITHIN THE PORT
INDUSTRIAL WORLD
WE'RE BECOMING THE NEW WESTSIDE
HISTORIC HISPANIC
SHINY CHROME RINGS AND GLASS BUILDINGS HOVER OVER THE LUXURY BOUTIQUES
PICTURE-PERFECT SUBURBIA. DON'T GET A SPEEDING TICKET! WHERE STRIP MALLS HAVE SMALL TOWN CHARM
LOW-RISE HOMES, GENTLE CURVING ROADS
MEDICAL WORLD
RODEOS, FOOTBALL CONVENTIONS, AND THE OLD 8th WONDER OF THE WORLD
IF ONLY THERE WERE A LITTLE MORE TREES!
"CHOPPED AND SCREWED" RAP LIVES HERE
RAILROAD TRACKS
LIVING QUIETLY ALONG THE BAYOU
610
10
90
45
59
288
<URBANE/>
www.mapurbane.com

Selection criteria



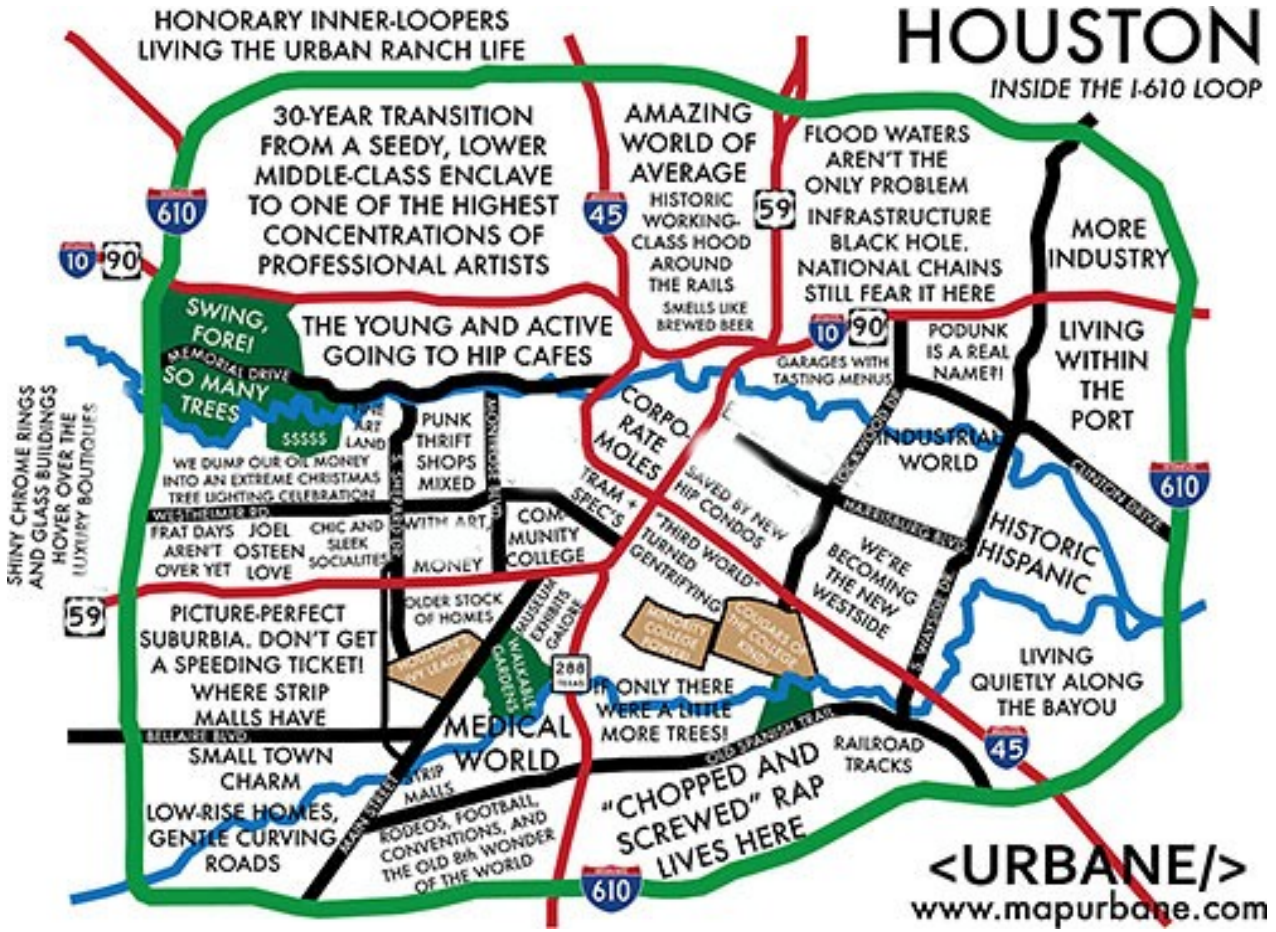
Selection criteria

- Near work



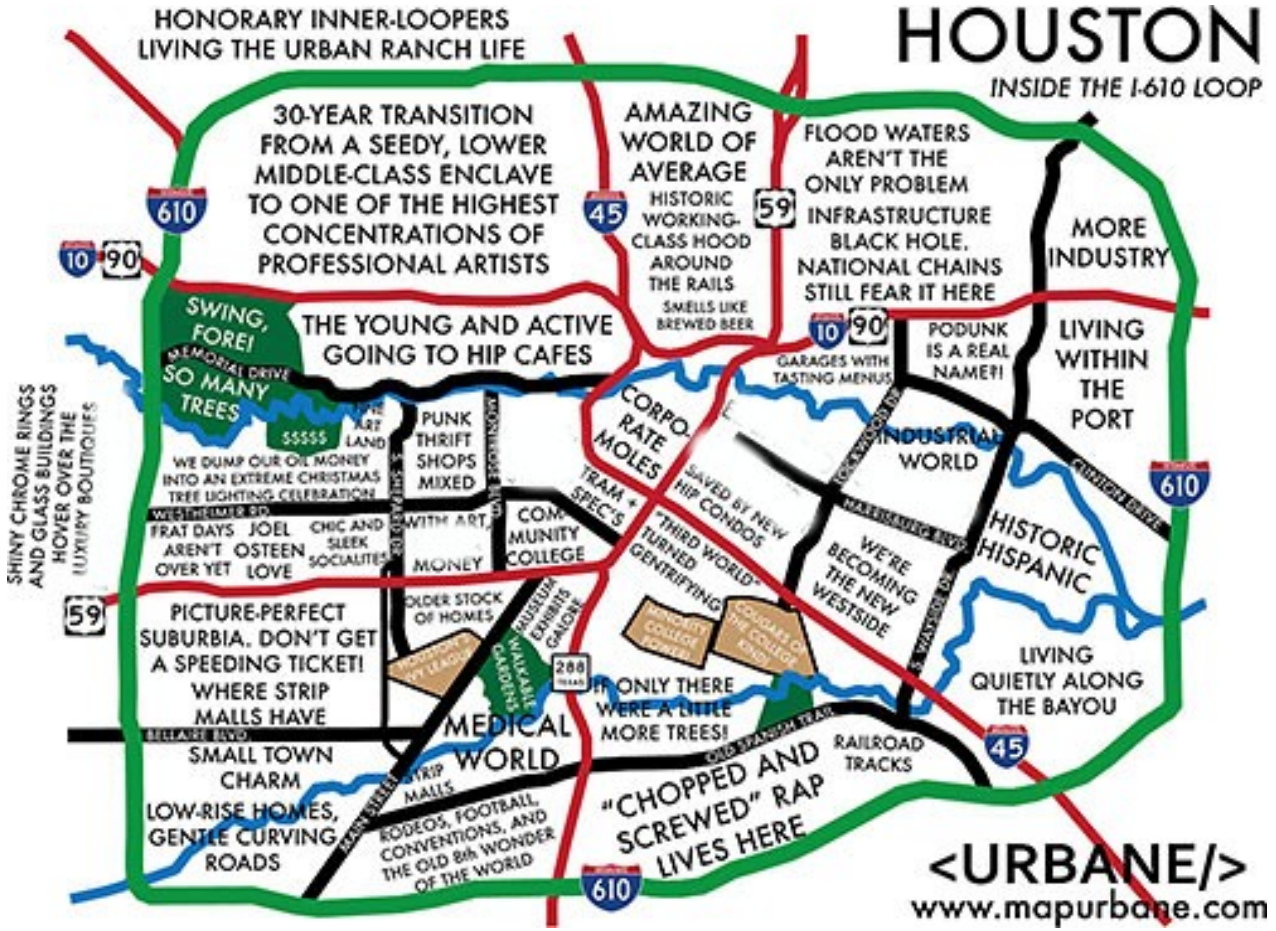
Selection criteria

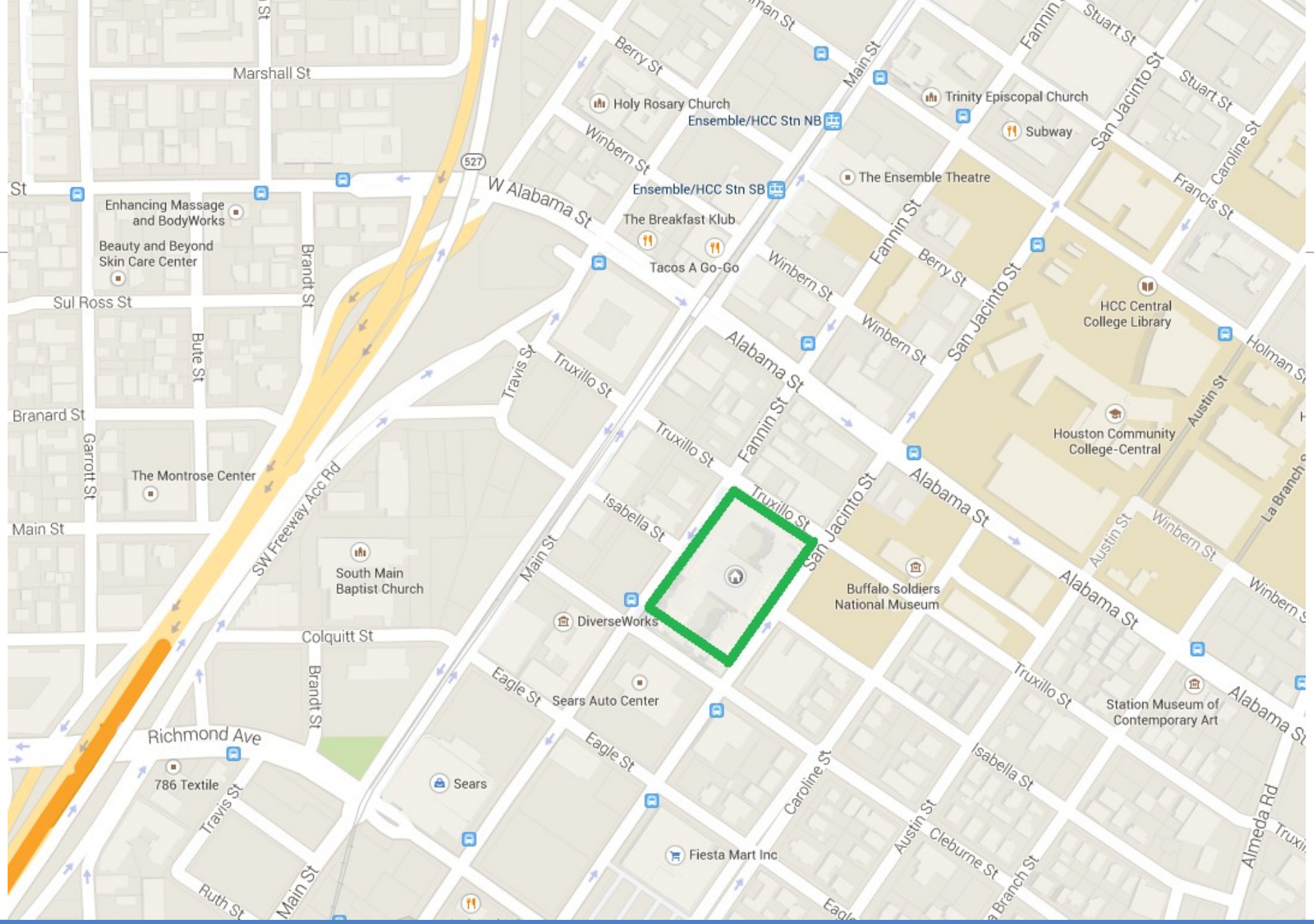
- Near work
- Near school



Selection criteria

- Near work
- Near school
- Near METRO

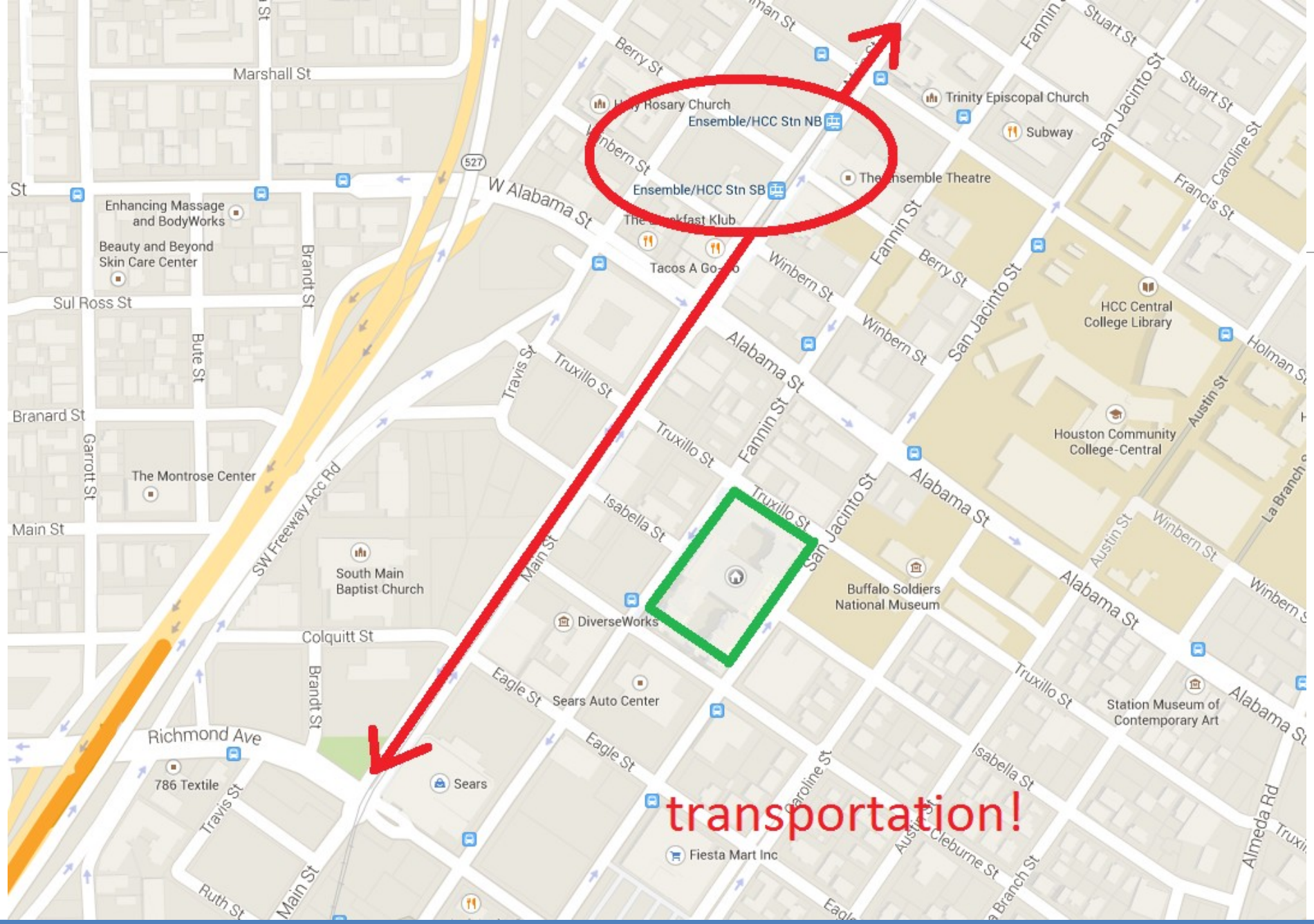


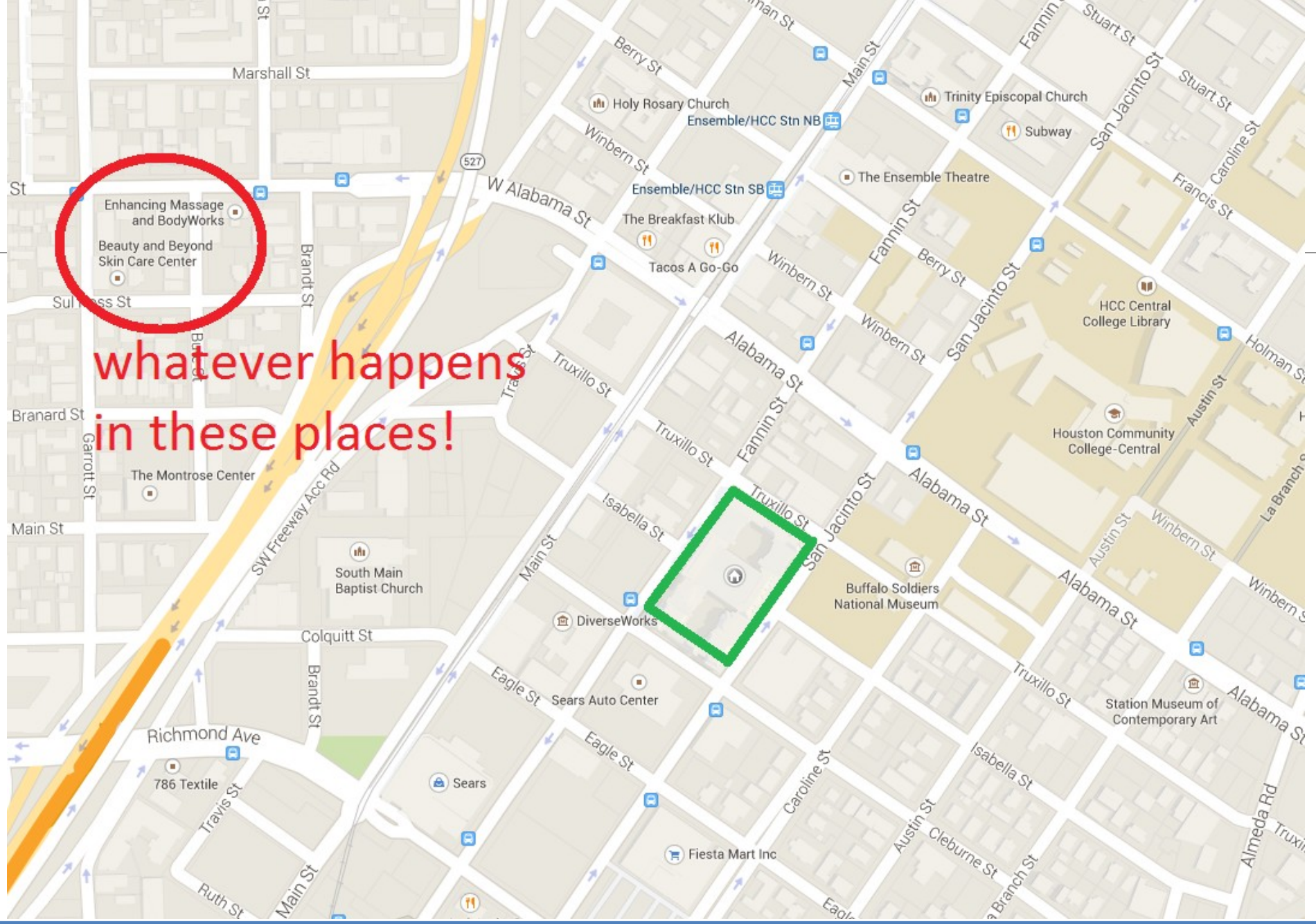

Marshall St
Berry St
Holy Rosary Church
Ensemble/HCC Stn NB
Main St
Trinity Episcopal Church
Subway
Stuart St
San Jacinto St
Caroline St
Winbern St
527
W Alabama St
Ensemble/HCC Stn SB
The Ensemble Theatre
Francis St
Enhancing Massage and BodyWorks
The Breakfast Klub
Beauty and Beyond Skin Care Center
Tacos A Go-Go
Fannin St
Brandt St
Sul Ross St
HCC Central College Library
Bute St
Travis St
Truxillo St
Alabama St
Holman St
Branard St
Houston Community College-Central
Austin St
Garrott St
The Montrose Center
SW Freeway Acc Rd
Isabella St
Main St
San Jacinto St
Buffalo Soldiers National Museum
South Main Baptist Church
DiverseWorks
Colquitt St
Eagle St
Sears Auto Center
Station Museum of Contemporary Art
Richmond Ave
786 Textile
Sears
Fiesta Mart Inc
Cleburne St
Almeda Rd
Ruth St

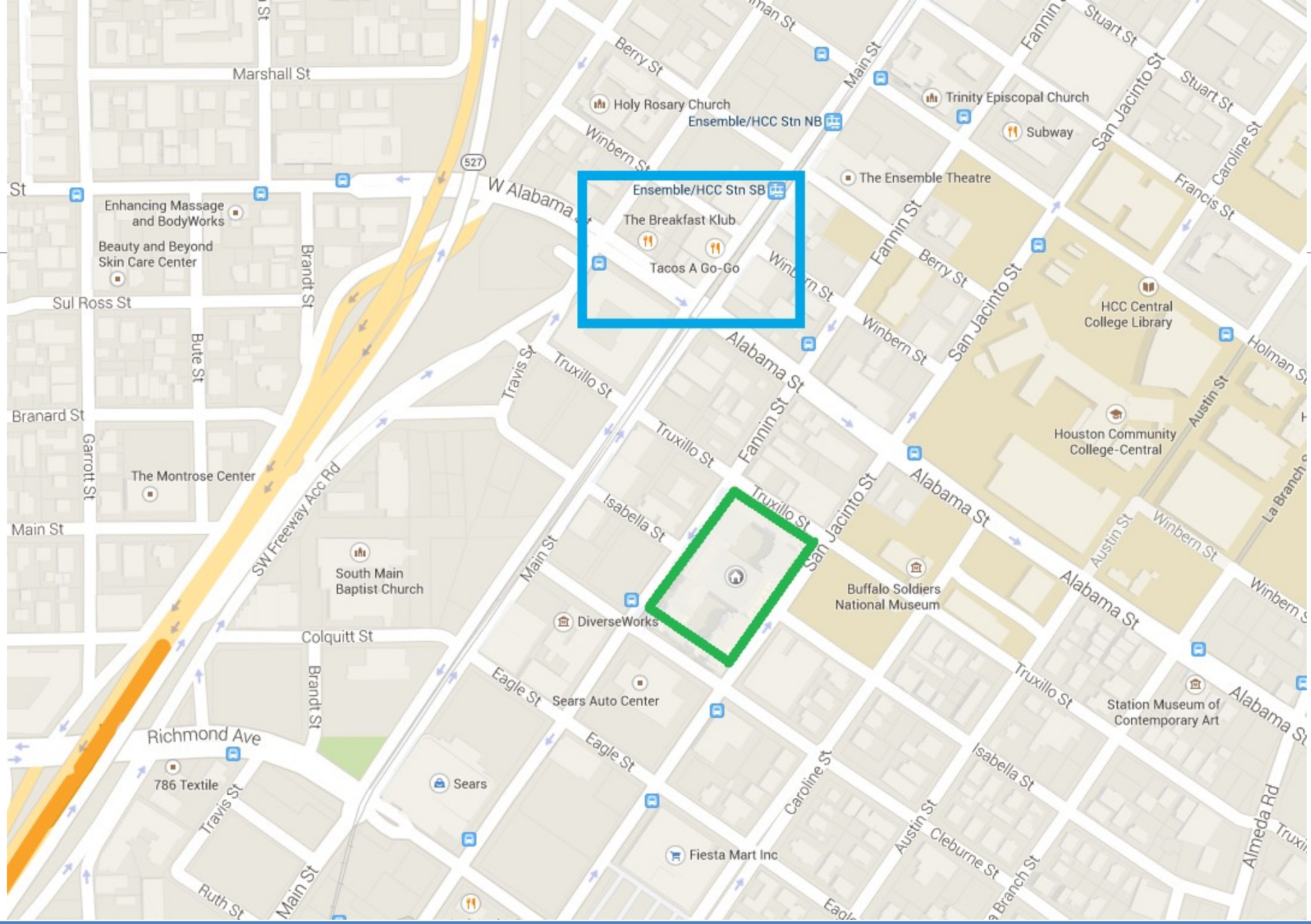

Marshall St
Holy Rosary Church
Ensemble/HCC Stn NB
Trinity Episcopal Church
Subway
Berry St
Main St
Fannin St
Stuart St
San Jacinto St
Caroline St
Winbern St
527
W Alabama St
Ensemble/HCC Stn SB
The Ensemble Theatre
Francis St
Enhancing Massage and BodyWorks
Beauty and Beyond Skin Care Center
The Breakfast Klub
Tacos A Go-Go
Brandt St
Sul Ross St
HCC Central College Library
Bute St
Travis St
Truxillo St
Alabama St
Holman St
Branard St
Houston Community College-Central
Austin St
Garrott St
The Montrose Center
SW Freeway Acc Rd
Isabella St
Main St
South Main Baptist Church
Buffalo Soldiers National Museum
DiverseWorks
Colquitt St
Eagle St
Sears Auto Center
Station Museum of Contemporary Art
Richmond Ave
786 Textile
Sears
Fiesta Mart Inc
Cleburne St
Almeda Rd
Ruth St
art!

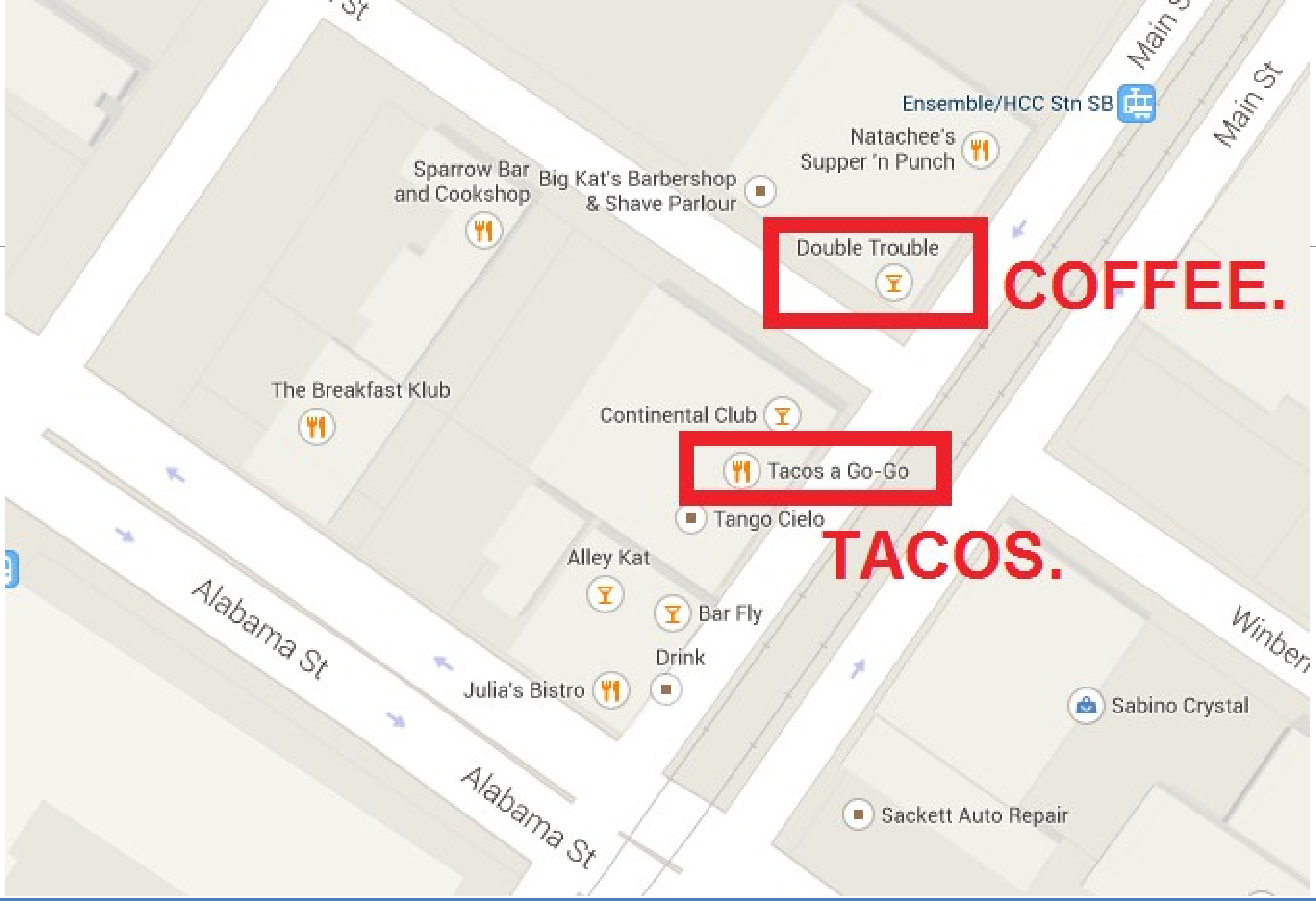

Holy Rosary Church
Trinity Episcopal Church
South Main Baptist Church
churches!
Marshall St
Berry St
Main St
Winbern St
W Alabama St
Ensemble/HCC Stn SB
Ensemble/HCC Stn NB
The Ensemble Theatre
The Breakfast Klub
Tacos A Go-Go
Enhancing Massage and BodyWorks
Beauty and Beyond Skin Care Center
Sul Ross St
Brandt St
Bute St
Branard St
Garrott St
The Montrose Center
Main St
SW Freeway Acc Rd
Travis St
Truxillo St
Fannin St
Alabama St
San Jacinto St
Isabella St
DiverseWorks
Colquitt St
Eagle St
Sears Auto Center
Richmond Ave
786 Textile
Sears
Fiesta Mart Inc
Ruth St
Cleburne St
Austin St
Caroline St
Francis St
Stuart St
HCC Central College Library
Houston Community College-Central
Buffalo Soldiers National Museum
Station Museum of Contemporary Art
Almeda Rd
Holman St
Subway


Marshall St
Holy Rosary Church
Ensemble/HCC Stn NB
Trinity Episcopal Church
Subway
Berry St
Main St
Stuart St
San Jacinto St
Caroline St
Winbern St
527
W Alabama St
Ensemble/HCC Stn SB
The Ensemble Theatre
Francis St
Enhancing Massage and BodyWorks
Beauty and Beyond Skin Care Center
The Breakfast Klub
Tacos A Go-Go
Fannin St
Brandt St
Sul Ross St
HCC Central College Library
Bute St
Travis St
Truxillo St
Alabama St
Holman St
Branard St
Houston Community College-Central
Austin St
Garrott St
The Montrose Center
SW Freeway Acc Rd
Isabella St
Main St
South Main Baptist Church
Buffalo Soldiers National Museum
DiverseWorks
Colquitt St
Eagle St
Sears Auto Center
Station Museum of Contemporary Art
Richmond Ave
786 Textile
Sears
Fiesta Mart Inc
Cleburne St
Almeda Rd
Ruth St
groceries!


Marshall St
Berry St
Holy Rosary Church
Ensemble/HCC Stn NB
Trinity Episcopal Church
Subway
Stuart St
San Jacinto St
Caroline St
Francis St
Winbern St
W Alabama St
Ensemble/HCC Stn SB
Fannin St
Enhancing Massage and BodyWorks
Beauty and Beyond Skin Care Center
Brandt St
Sul Ross St
Bute St
Tacos A Go-Go
HCC Central College Library
Holman St
Austin St
Houston Community College-Central
Travis St
Truxillo St
Alabama St
Branard St
Garrott St
The Montrose Center
Main St
SW Freeway Acc Rd
Isabella St
South Main Baptist Church
DiverseWorks
Buffalo Soldiers National Museum
Colquitt St
Eagle St
Sears Auto Center
Station Museum of Contemporary Art
Richmond Ave
786 Textile
Sears
Fiesta Mart Inc
Cleburne St
Almeda Rd
Ruth St
transportation!


whatever happens
in these places!
Marshall St
Berry St
Holy Rosary Church
Ensemble/HCC Stn NB
Trinity Episcopal Church
Subway
Winbern St
W Alabama St
Ensemble/HCC Stn SB
The Ensemble Theatre
The Breakfast Klub
Tacos A Go-Go
Enhancing Massage and BodyWorks
Beauty and Beyond Skin Care Center
Brandt St
Fannin St
San Jacinto St
Stuart St
Caroline St
Francis St
HCC Central College Library
Holman St
Houston Community College-Central
Austin St
Alabama St
Truxillo St
Isabella St
Branard St
Garrott St
The Montrose Center
Main St
SW Freeway Acc Rd
South Main Baptist Church
DiverseWorks
Buffalo Soldiers National Museum
Colquitt St
Eagle St
Sears Auto Center
Richmond Ave
786 Textile
Travis St
Sears
Fiesta Mart Inc
Ruth St
Cleburne St
Almeda Rd
Station Museum of Contemporary Art
La Branch St


Marshall St
Berry St
Holy Rosary Church
Ensemble/HCC Stn NB
Main St
Trinity Episcopal Church
Subway
Stuart St
San Jacinto St
Caroline St
Winbern St
527
W Alabama St
Ensemble/HCC Stn SB
The Breakfast Klub
Tacos A Go-Go
The Ensemble Theatre
Francis St
Fannin St
Enhancing Massage and BodyWorks
Beauty and Beyond Skin Care Center
Brandt St
Sul Ross St
Bute St
HCC Central College Library
Holman St
Alabama St
Travis St
Truxillo St
Branard St
Garrott St
The Montrose Center
Houston Community College-Central
Austin St
SW Freeway Acc Rd
Isabella St
Main St
South Main Baptist Church
Buffalo Soldiers National Museum
DiverseWorks
Colquitt St
Eagle St
Sears Auto Center
Station Museum of Contemporary Art
Richmond Ave
786 Textile
Sears
Fiesta Mart Inc
Cleburne St
Almeda Rd
Ruth St


Ensemble/HCC Stn SB
Natachee's
Supper 'n Punch
Sparrow Bar
and Cookshop
Big Kat's Barbershop
& Shave Parlour
Double Trouble
COFFEE.
The Breakfast Klub
Continental Club
Tacos a Go-Go
Tango Cielo
TACOS.
Alley Kat
Bar Fly
Drink
Julia's Bistro
Sabino Crystal
Sackett Auto Repair
Alabama St
Alabama St
Main St
Main St

“If you park your car outside, make sure only to park it on Fannin or Truxillo.”

– Security Guard

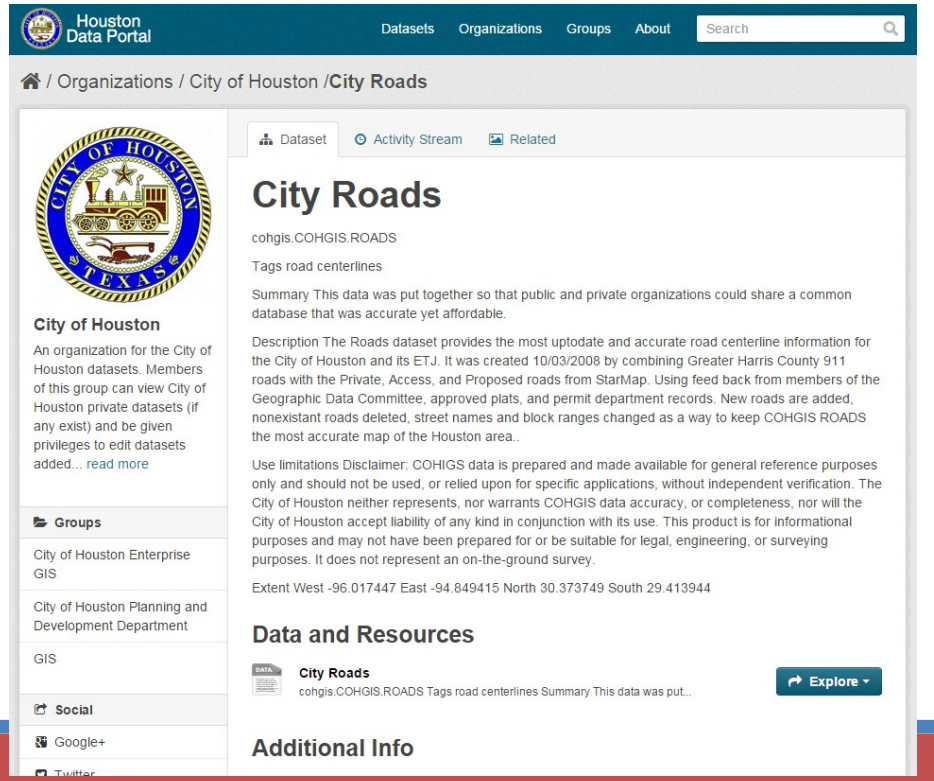
???

How can one side of a building be more secure than another side?

Obtaining the Data

Houston Police Department's Neighborhood (Police Beat) Crime Statistics
- http://www.houstontx.gov/police/cs/beatpages/beat_stats.htm

Also used the City of Houston's "ROADS" shapefile, via the Open Data Portal
- http://data.ohouston.org/dataset/city-roads

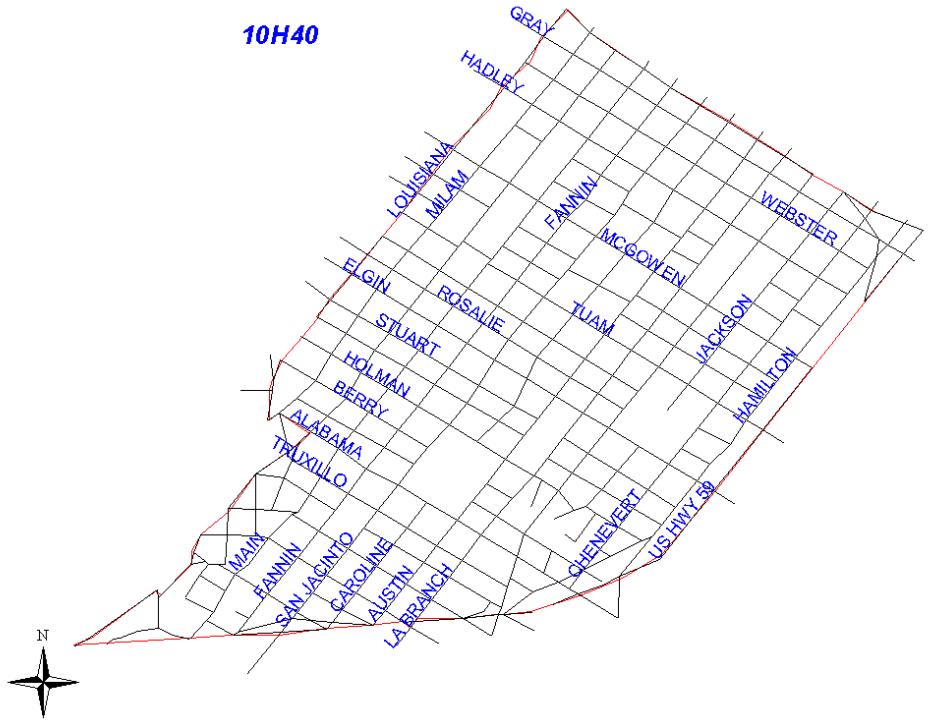


Data for 10H40

7,445 entries since June 2009
Averages out to ~108 crimes per month (3-4 per day)

Zip codes included:

- 77002
- 77003
- 77004
- 77006



Attribute data per crime

Date	Hour	Offense Type	Beat	Premise	BlockRange	StreetName	Type	Suffix	# offenses
2/19/2015	22	Theft	10H40	Bus Station	2100-2199	MAIN	ST	-	1
2/20/2015	02	Theft	10H40	Road, Street, or Sidewalk	2300-2399	SAN JACINTO	-	-	1
2/25/2015	09	Theft	10H40	Grocery Store or Supermarket	4200-4299	SAN JACINTO	-	-	1
2/27/2015	07	Auto Theft	10H40	Apartment Parking Lot	2800-2899	AUSTIN	ST	-	1
2/27/2015	12	Theft	10H40	Bar or Night Club Parking Lot	2400-2499	SAN JACINTO	-	-	1
2/28/2015	02	Aggravated Assault	10H40	Road, Street, or Sidewalk	800-899	ROSALIE	-	-	1
1/19/2015	14	Theft	10H40	Bar or Night Club	2900-2999	TRAVIS	-	-	1
2/6/2015	03	Theft	10H40	Bar or Night Club	2100-2199	MAIN	ST	-	1
2/4/2015	06	Theft	10H40	Bus Stop	2000-2099	MAIN	-	-	1
2/8/2015	21	Theft	10H40	Road, Street, or Sidewalk	1100-1199	DREW	ST	-	1
2/11/2015	11	Burglary	10H40	Office Building	2000-2099	MAIN	ST	-	1
2/11/2015	13	Theft	10H40	Commercial Parking Lot or Garage	1000-1099	FRANCIS	ST	-	1
2/11/2015	22	Theft	10H40	Apartment Parking Lot	2500-2599	MILAM	-	-	1
2/1/2015	18	Theft	10H40	Apartment Parking Lot	2700-2799	TRAVIS	-	-	1
1/23/2015	16	Theft	10H40	Bar or Night Club	2000-2099	MAIN	ST	-	1
2/1/2015	01	Theft	10H40	Bar or Night Club	2400-2499	SAN	-	-	1

Date
Hour (0 – midnight, 23 – 11:00pm)

Offense Type
- Murder
- Rape
- Robbery
- Aggravated Assault
- Burglary
- Theft
- Auto Theft

Beat
Premise Code
Block Range
Street Name
# of Offenses

Attribute data per crime

Premise (Location) Codes

Premise Type	Premise Description
01A	AIRPORT TERMINAL
01B	BUS STATION
01K	RAILROAD TRACK/RIGHT OF WAY
01P	PARK & RIDE TERMINAL
01R	LIGHT RAIL VEHICLE
01T	TRAIN TERMINAL
02B	BANK
02C	CREDIT UNION
02S	SAVINGS AND LOAN INSTITUTIONS
02V	VACANT BANK
03B	BAR/NIGHT CLUB
03S	SEXUALLY ORIENTED CLUB
040	CHURCH/SYNAGOGUE/TEMPLE
04V	VACANT CHURCH/SYNAGOGUE/TEMPLE
05A	AMUSE. PARK,BOWL. ALLEY,SKATE RINK
05B	BARBER AND BEAUTY SHOPS
05C	COMMERCIAL BUILDING
05D	CAR WASH
05E	AUTO REPAIR
05F	FACTORY/MANUFACTURING/INDUSTRIAL
05G	GYM,RECREAT,CLUB HSE,INDR POOL,SPA
05H	BODY SHOP
05L	LAUNDRY/DRY CLEANERS/WASHATERIAS
05M	MALL COMMON AREA
05N	MAINTENANCE/BUILDING SERVICES
05O	OFFICE BUILDING
05P	POOL HALL/GAME ROOM
05Q	CHECK CASHING PLACES
05R	APARTMENT/RENTAL OFFICE
05S	STADIUM/SPRTS ARENA/RACE TRACK
05T	THEATRES,DINNER THEATERS,AUDITOR.
05U	UTILITY COMPANY,ELECTRIC,GAS,WATER
05V	VACANT BUILDING (COMMERCIAL)
05W	WAREHOUSE
05X	VEHICLE/AUTO SALES/LEASE/AUTO PARTS STORE
05Y	VACANT INDUSTRIAL/MANUFACTURING/INDUSTRIAL
05Z	MISC. BUSINESS (NON-SPECIFIC)
060	CONSTRUCTION SITE
070	CONVENIENCE STORE
080	DEPARTMENT/DISCOUNT STORE
09D	DRUG STORE/MEDICAL SUPPLY
09H	HOSPITAL
09P	PHYSICIAN'S OFFICE
09R	REHABILITATION CENTER
09V	VACANT HOSPITAL
100	FIELD/WOODS
11C	CONVENTION CENTER/EXHIBIT HALLS
11F	FIRE STATION
11G	GOVERNMENT/PUBLIC BUILDING
11L	LIBRARIES, MUSEUMS
11P	POLICE STATION
11R	PARKS & RECREATION, ZOO, SWIM POOL
11S	SOCIAL SERVICES/PUBLIC CHARITIES
11V	VACANT GOVERNMENT/PUBLIC BUILDING
120	GROCERY/SUPERMARKET
12V	VACANT GROCERY/SUPERMARKET
13A	ALLEY
13B	BUS STOP
13C	CONTRA-FLOW/HOV
13H	HIGHWAY/FREEWAY
13R	ROAD/STREET/SIDEWALK
13S	FREEWAY SERVICE ROAD
13T	TOLLWAY
140	HOTEL/MOTEL/ETC.
14V	VACANT HOTEL/MOTEL/ETC.
150	JAIL/PRISON
15V	VACANT JAIL/PRISON
160	LAKE/WATERWAY/BAYOU
170	LIQUOR STORE
18A	APARTMENT PARKING LOT
18B	BANK/SAVING INSTITUTION PARKING LOT
18C	CONVENIENCE STORE PARKING LOT
18D	LAUNDRY/DRY CLEANERS PARKING LOT
18G	GROCERY/SUPERMARKET PARKING LOT
18H	HOSPITAL PARKING LOT
18L	LIQUOR STORE PARKING LOT
18M	HOTEL/MOTEL PARKING LOT
18N	BAR/NIGHT CLUB PARKING LOT
18O	OTHER PARKING LOT
18P	COMMERCIAL PARKING LOT/GARAGE
18R	RESTAURANT/CAFETERIA PARKING LOT
18S	SEXUALLY ORIENTED BUSINESS PARKING LOT
18T	STRIP BUSINESS CENTER PARKING LOT
18U	MALL PARKING LOT
18W	CHURCH/SYNAGOGUE/TEMPLE PARKING LOT
190	RENTAL STORAGE FACILITY
19V	VACANT STORAGE FACILITY (BARNS,GARAGES,WAREHOUSES,ETC.)
20A	APARTMENT
20C	CONDOMINIUM
20D	DRIVEWAY
20G	GARAGE/CARPORT
20H	HIGH RISE
20L	APARTMENT LAUNDRY
20M	MOBILE HOME
20N	NURSING HOME
20P	MULTI-PLEX HME(DUPLEX,TRIPLEX ETC)
20R	RESIDENCE/HOUSE
20V	VACANT SINGLE OCCUPANCY RESIDENCE (HOUSES,TOWNHOUSES,DUPLEXES, ETC.)
20W	VACANT OTHER RESIDENTIAL (APARTMENT,INN,DORMS,BOARDING HOUSE)
210	RESTAURANT/CAFETERIA
21V	VACANT RESTAURANT
22C	COMMERCIAL OR TRAINING SCHOOL
22D	DAYCARE/CHILD CARE/KINDERGARTEN
22E	ELEMENTARY SCHOOL
22H	HIGH SCHOOL
22M	MIDDLE SCHOOL
22P	PRIVATE SCHOOL
22U	UNIVERSITY/COLLEGE
22V	VACANT SCHOOL/COLLEGE
23S	SERVICE/GAS STATION
240	SPECIALTY STORE (NON-SPECIFIC)
24A	ADULT BOOK STORE/NEWSSTAND
24B	BOOK,RECORD,STATIONARY,OFFICE SUP.
24C	CLOTHING STORE
24E	ELECTRONICS STORE, ELECTRICAL SUP.
24F	FURNITURE, APPLIANCES, RADIOS, TV
24G	GARDEN SUPPLY, NURSERY, FLORIST

Date
Hour (0 – midnight, 23 – 11:00pm)
Offense Type
- Murder
- Rape
- Robbery
- Aggravated Assault
- Burglary
- Theft
- Auto Theft

Beat
Premise Code
Block Range
Street Name
# of Offenses

Potential problems with the data

Not sure if “date” refers to “date the complaint was filed”, or “date the event occurred”.

Same for “hour” – is this the hour the crime was reported? Or the hour the crime occurred?

Two months were missing from the data set; some previous months’ data is intermixed

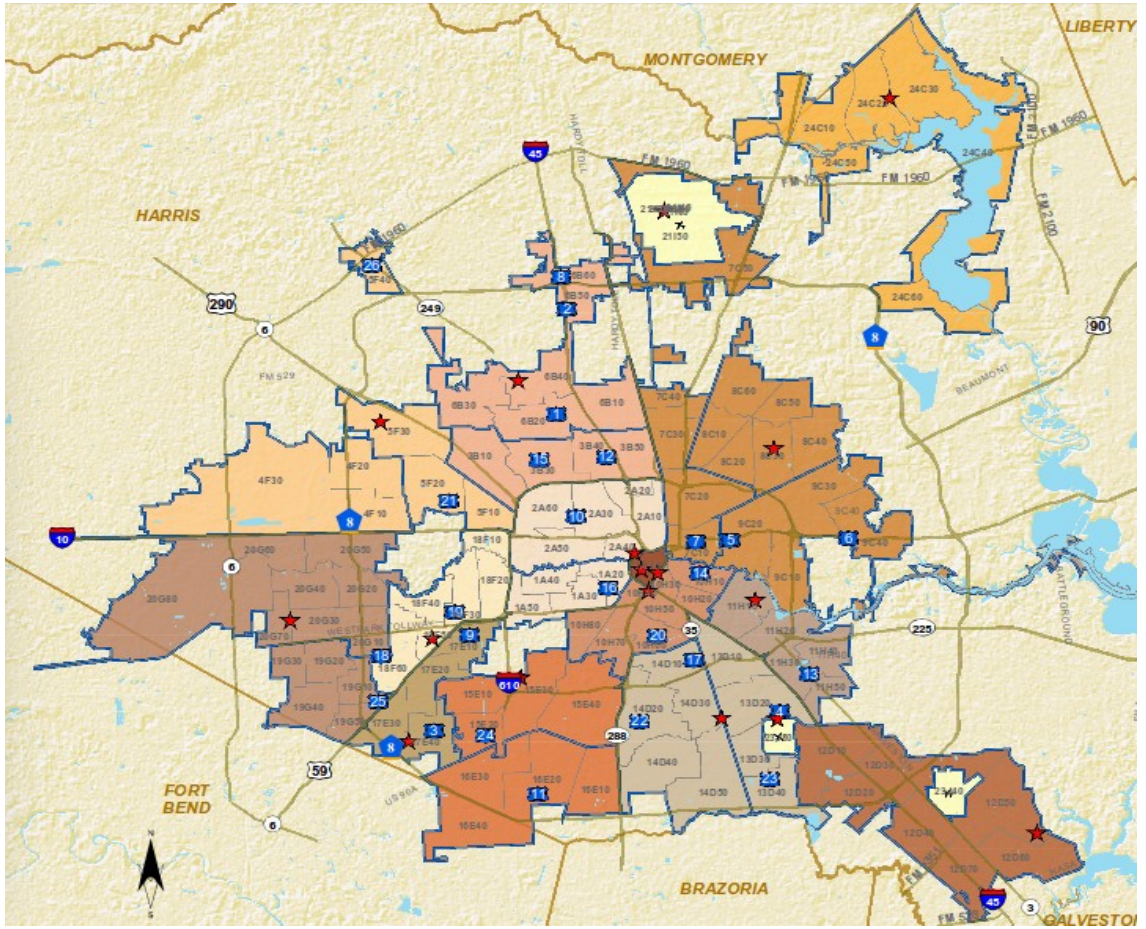
“UNK” values for block ranges occurred several times.

Table changed format some time in 2010 (easily dealt with using Beautiful Soup; meant hand-manipulation for initial analysis).

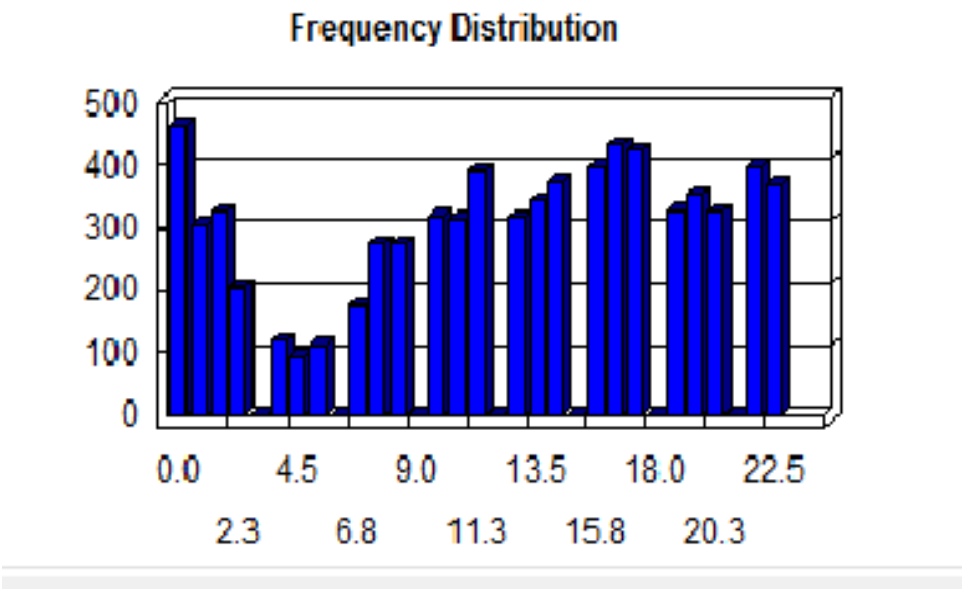
What are the most common crimes committed?

10H40 seems to get everything…

- Aggravated Assault – 511
- Auto Theft – 427
- Burglary – 554
- Murder – 4
- Rape – 45
- Robbery – 632
- **Theft – 5,272**



When are those crimes most likely to happen?



Most occur around midnight and between 4:00 – 6:00pm.

The least number of crimes occur between 3:00am and 6:00am.

Not quite sure why there would be such a spike between 4:00 – 6:00pm – maybe because so many people are headed home from work and away from school?

Where are crimes most likely to happen?

Merged the “BLOCK RANGE” and “STREET NAME” attributes in order to make a single attribute called “Block_Street”

Python:

```
str([block_range]) + “ “ + ([street_name]).upper()
```

Top instances
(>100 crimes committed):

- 4200 – 4299 San Jacinto: 864
- 4200 – 4299 Main St: 566
- 2100 – 2199 Main St: 245
- 2400 – 2499 San Jacinto: 224
- 1000 – 1099 Elgin St: 147
- 2900 – 2999 Travis St: 114
- 4600 – 4699 Main St: 101

Table

Sum_Output

OID	Block_Street	Count_Block_Street
419	4200-4299 SAN JACINTO	864
418	4200-4299 MAIN	566
197	2100-2199 MAIN	245
235	2400-2499 SAN JACINTO	224
14	1000-1099 ELGIN	147
294	2900-2999 TRAVIS	114
434	4600-4699 MAIN	101
193	2100-2199 FANNIN	99
187	2000-2099 MAIN	92
270	2700-2799 TRAVIS	82
347	3400-3499 MILAM	76
343	3400-3499 FANNIN	74
259	2600-2699 TRAVIS	72
6	1000-1099 ALABAMA	69
209	2200-2299 MAIN	67
241	2500-2599 FANNIN	57
146	1700-1799 GRAY	51
81	1300-1399 HOLMAN	48
250	2600-2699 CAROLINE	48
268	2700-2799 MILAM	48
400	4000-4099 FANNIN	48
506	MAIN ST	47
432	4500-4599 MAIN	45
184	2000-2099 CRAWFORD	44

So how close is that to me?

Three of the most high-crime areas are located fairly close to my apartment complex:

- 4200 – 4299 San Jacinto St - 864
- 4200 – 4299 Main St – 566
- 4600 – 4699 Main St – 101

20.6%

of crimes in the 10H40 police beat

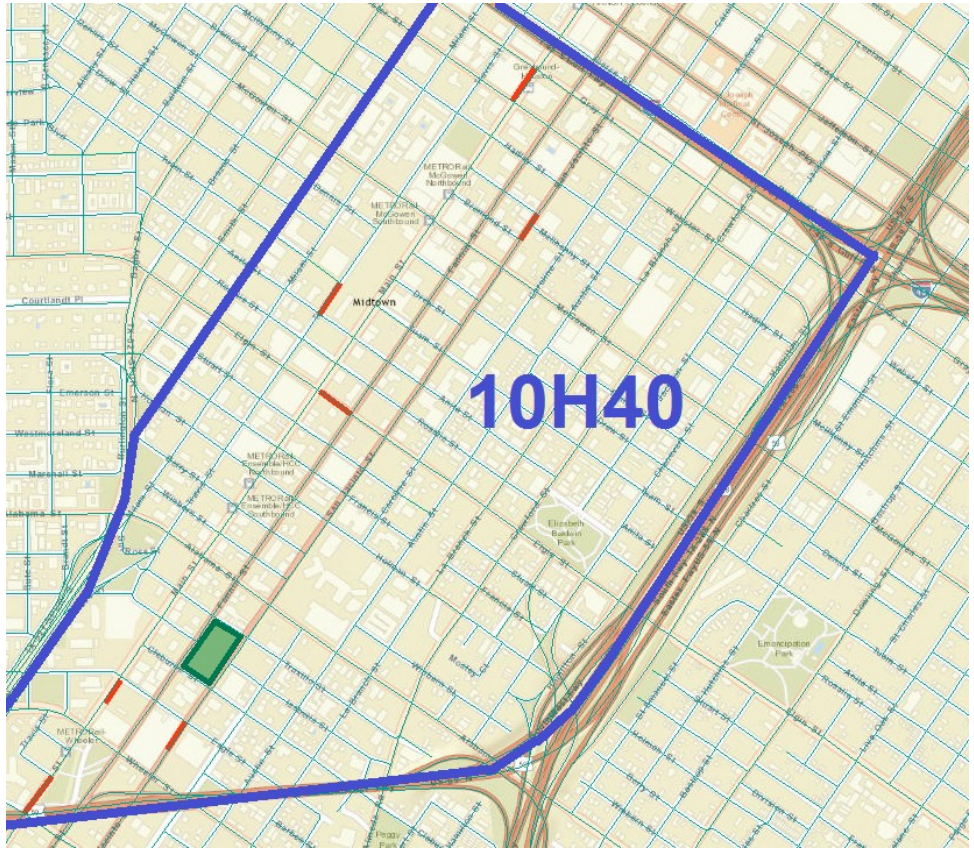
Which side of the Ventana is most dangerous? Least dangerous?

- 3900 – 4000 Fannin Street
- 3900 – 4000 San Jacinto Street
- 1000 – 1200 Truxillo Street
- 1000 – 1200 Cleburne Street

Most dangerous:
San Jacinto Street and Cleburne Street (like the guard said)

Least dangerous:
Fannin Street and Truxillo Street

Police Beat 10H40



Most of the crime appears to be occurring along the MetroRail

The Greyhound Station appears to have a very high crime rate, as well.

Of the three MetroRail stops in my vicinity, the HCC/Ensemble Theatre stop has a lower crime rate than the Wheeler Station stop or the McGowen Street station stop.

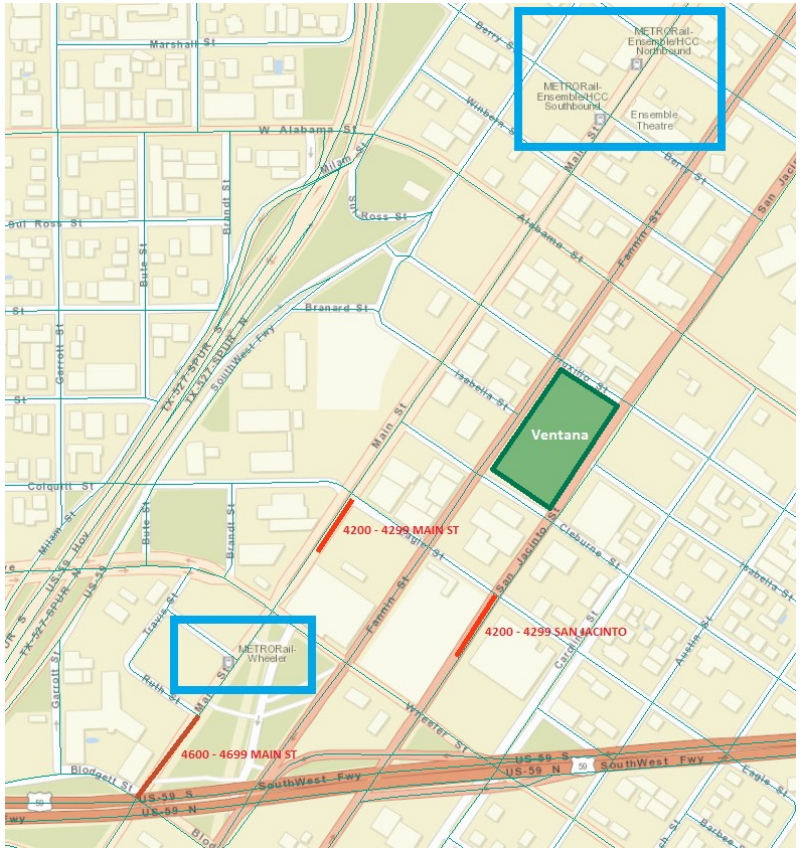
Which MetroRail stop is the safest?

There are two MetroRail options for me when I go to work or to school:

- Wheeler Station
- Ensemble / HCC Station

Both are equidistant from my apartment

It’s obvious to see, when mapped, that the safest MetroRail station for me to go to – no matter what time of day – is the HCC / Ensemble Theatre Station.



Which route is the safest to walk to work? When is the safest time?

Route to work (as recommended by Google Maps transportation routing):

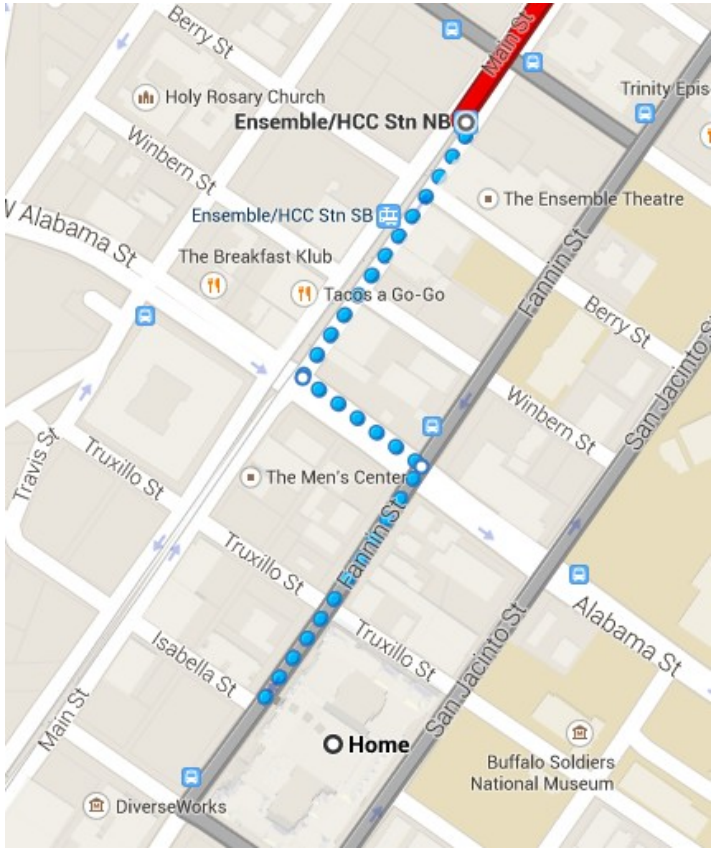
- 4000 Fannin to 3800 Fannin
- 1100 Alabama to 1000 Alabama
- 3700 Main St to 3500 Main St

Crimes along this route:
160 total (2.15%)

Weighted route using ArcGIS Network Analyst tool – instead of distance, optimizes for route with least amount of criminal activity for given time of day

Safest Route – **104 crimes total (1.4%)**

- 4000 Fannin St to 3600 Fannin St
- 1100 Berry St to HCC Station



Which route is the safest to walk to work? When is the safest time?

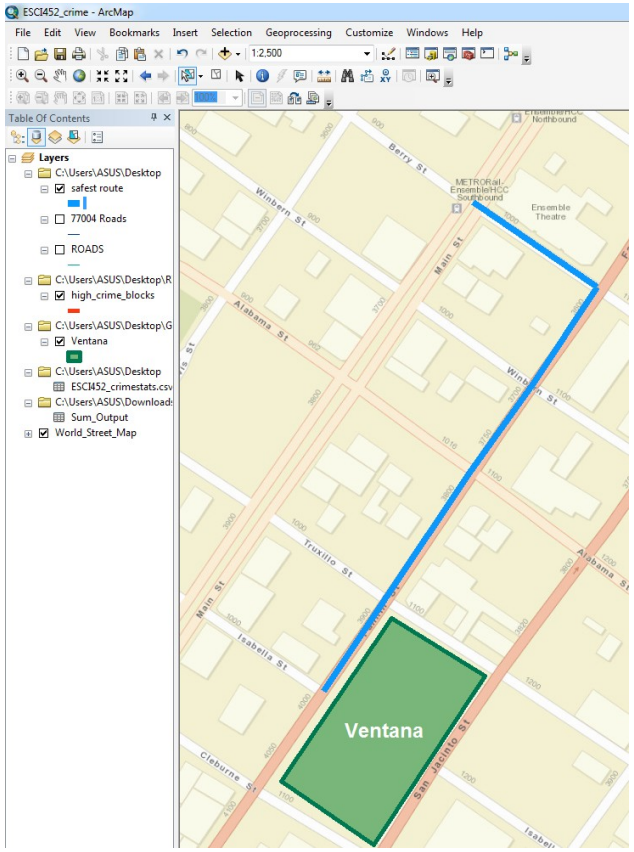
The most safe time to walk along the safest route is between 3:00am and 5:00am

Zero crimes have been committed along the Fannin Street route since 2009 for that time frame

6 thefts, 2 burglaries, and 1 aggravated assault have occurred between the hours of 7:00am – 9:00am along the Fannin Street route

When I normally walk to work, between 5:00 and 7:00 am, there has been 1 rape, 2 auto thefts, and 1 theft.

4 out of 7,445 instances = not that bad.



How much was risk reduced?

The route I'd currently been walking to the MetroRail was to the Wheeler Station stop

Three of the most dangerous blocks in Houston

Total number of offenses between the hours of 5:00 and 7:00am since June 2009:

72

~1% of the total offenses for 10H40

Thefts, robberies, rapes, aggravated assaults

Risk was reduced to .0537%

“It would be great if you could do this for all of Houston! I worry so much about biking home from school to my apartment...”

“Could you do this for the Westheimer area? Never know where to park my car...”

“Could you do this for the bike paths along the Bayou?”

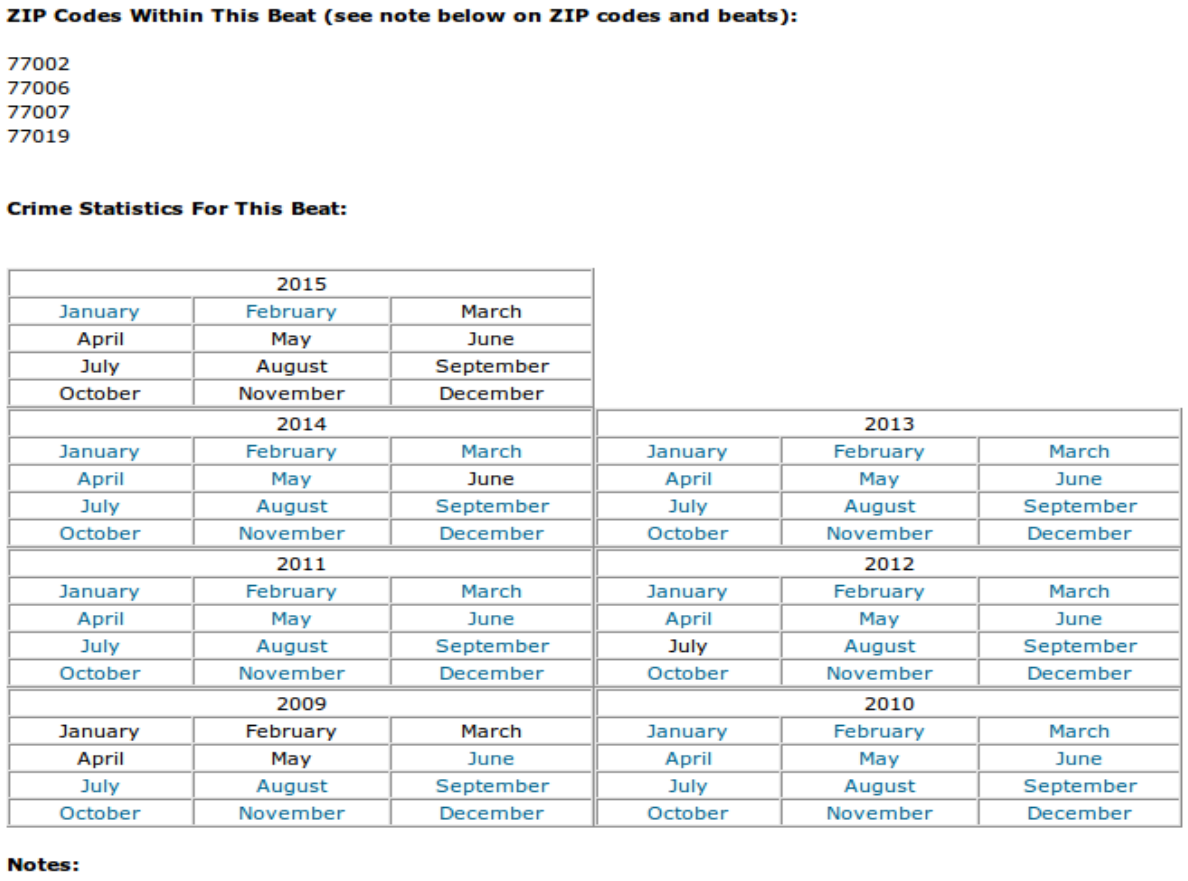
Absolutely!

Absolutely!

...but there's no way in heck I'm doing that by hand.

Data format on HPD's web site

- 92 police beats
- 69 months of data per police beat
- 6,348 separate .csv files



ZIP Codes Within This Beat (see note below on ZIP codes and beats):

77002
77006
77007
77019

Crime Statistics For This Beat:

2015		
January	February	March
April	May	June
July	August	September
October	November	December

2014			2013		
January	February	March	January	February	March
April	May	June	April	May	June
July	August	September	July	August	September
October	November	December	October	November	December

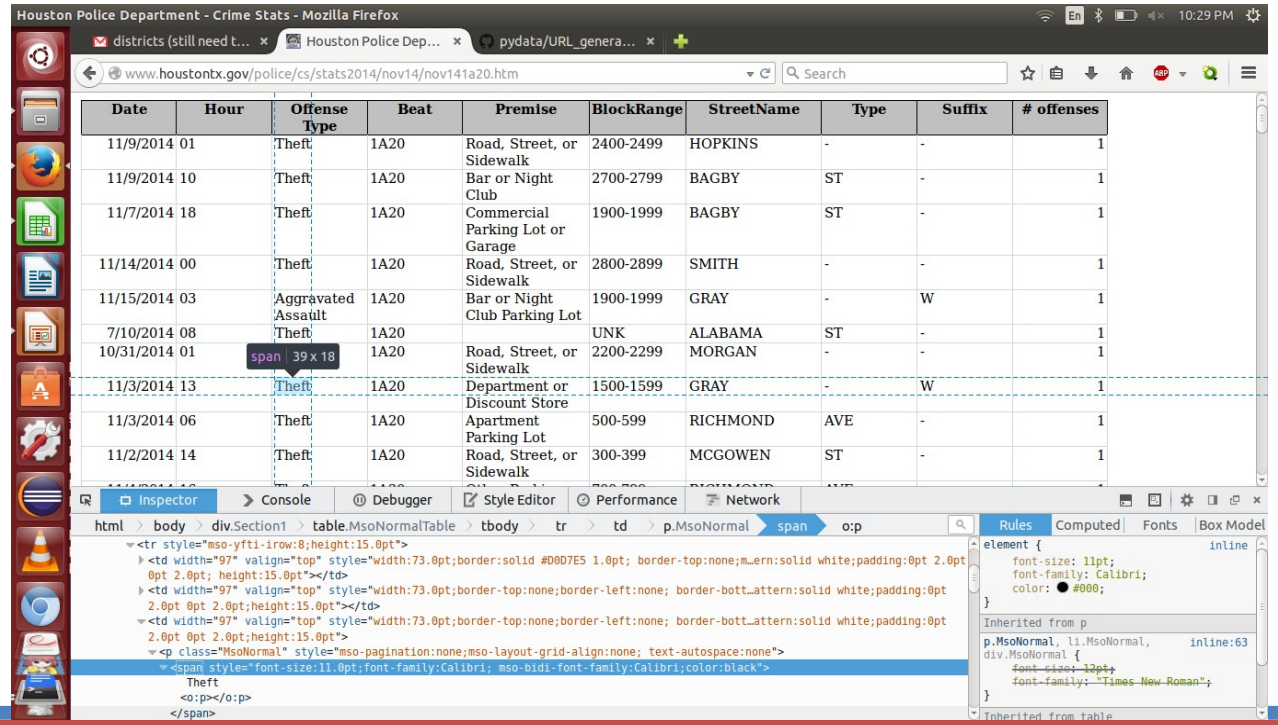
2011			2012		
January	February	March	January	February	March
April	May	June	April	May	June
July	August	September	July	August	September
October	November	December	October	November	December

2009			2010		
January	February	March	January	February	March
April	May	June	April	May	June
July	August	September	July	August	September
October	November	December	October	November	December

Notes:

Suitable for automation?

- URL's for each police beat are in a consistent format
- Tables and associated data are also in a (fairly) consistent format
- Same tag `<span>` used for each cell value



Pain points

- Column order changed depending on date, beat
  (never the number of columns, though, or their names)
- Some months' data was unavailable
  (webpage doesn't exist, or its name had been misspelled)
- UTF-8 vs. Unicode; strings vs. integers
- Dates with dissimilar formats
- “UNK” values vs. empty strings
- City blocks' names not in same format as CoH's ROADS.shp
  (`str([block_range]) + “ “ + ([street_name]).upper()`)

Pain points – all solvable!

- Column order changed depending on date, beat
  (never the number of columns, though, or their names)
- Some months' data was unavailable
  (webpage doesn't exist, or its name had been misspelled)
- UTF-8 vs. Unicode; strings vs. integers
- Dates with dissimilar formats
- “UNK” values vs. empty strings
- City blocks' names not in same format as CoH's ROADS.shp
  (`str([block_range]) + “ “ + ([street_name]).upper()`)

Tools and methodology

```python
#! /usr/bin/python

import time
import datetime

HPD_police_beats = ["1A10", "1A20", "1A30", "1A40", "1A50", "3B10", "3B20", "3B30", "3B40", "3B50", "7C10", "7C20",

full_years = []
months_to_check = []
URLS = []
months = ["Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"]
now = datetime.datetime.now()

# Check's for the current year
# HPD only has police beat data for 2009+
for i in range(2010, now.year, 1):
        full_years.append(str(i))

# Checks for user's current month
for i in range(12):
        if now.strftime("%b") != months[i]:
                months_to_check.append(months[i])
        else:
                break

# Removes previous month from list of months to check in the current year
# Usually there's a 2-month delay in HPD's police beat data
months_to_check.pop()

def URL_populator():
        for i in  months[5:]:
                for j in HPD_police_beats:
```

- Created a script to generate a list of URL's for HPD police beats, using current date (-2 months)
- Libraries used: `datetime, time`

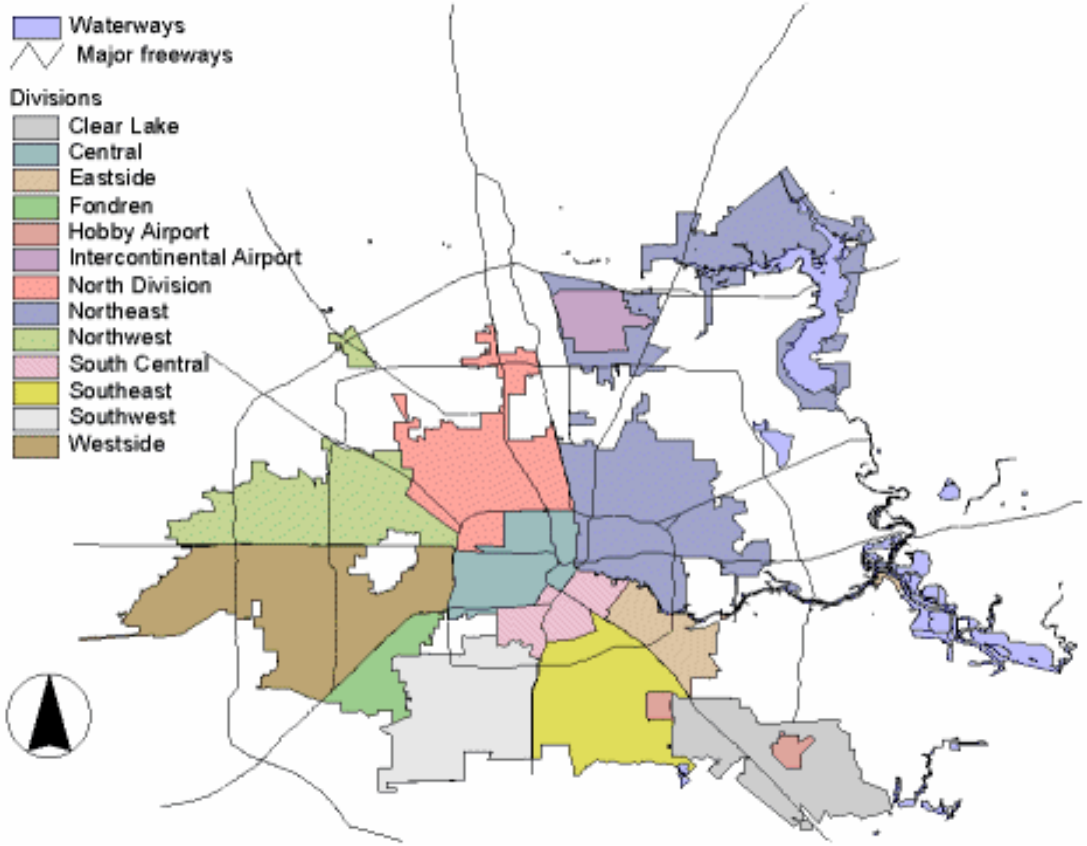
Tools and methodology

```
import urllib2
import pprint
import csv
import requests
import URL_generator as populator
from bs4 import BeautifulSoup


# next step: looping through list of URL's produced by URL_generator.py
urls = populator.URL_populator()

for url in urls:
        html = urllib2.urlopen(url).read()
        soup = BeautifulSoup(html)

        def chunks(l,n):
                for i in xrange(0, len(l), n):
                        yield l[i:i+n]

        for row in soup('table', {'class':'MsoNormalTable'}):
                holder = []
                values = row.find_all('span')
                for i in values:
                        try:
                                holder.append(str(i.get_text()))
                        except:
                                UnicodeDecodeError
                                holder.append("Unknown")
```

- Cycled through the list of URL's pulling everything from within the `<span>` tag
- `UnicodeDecodeError`
- `r = requests.get()`
  `r.raise_for_status()`
- `open('out.csv','a')`
- `BeautifulSoup(html)`
- `soup('table', {'class': 'MsoNormalTable'}):`

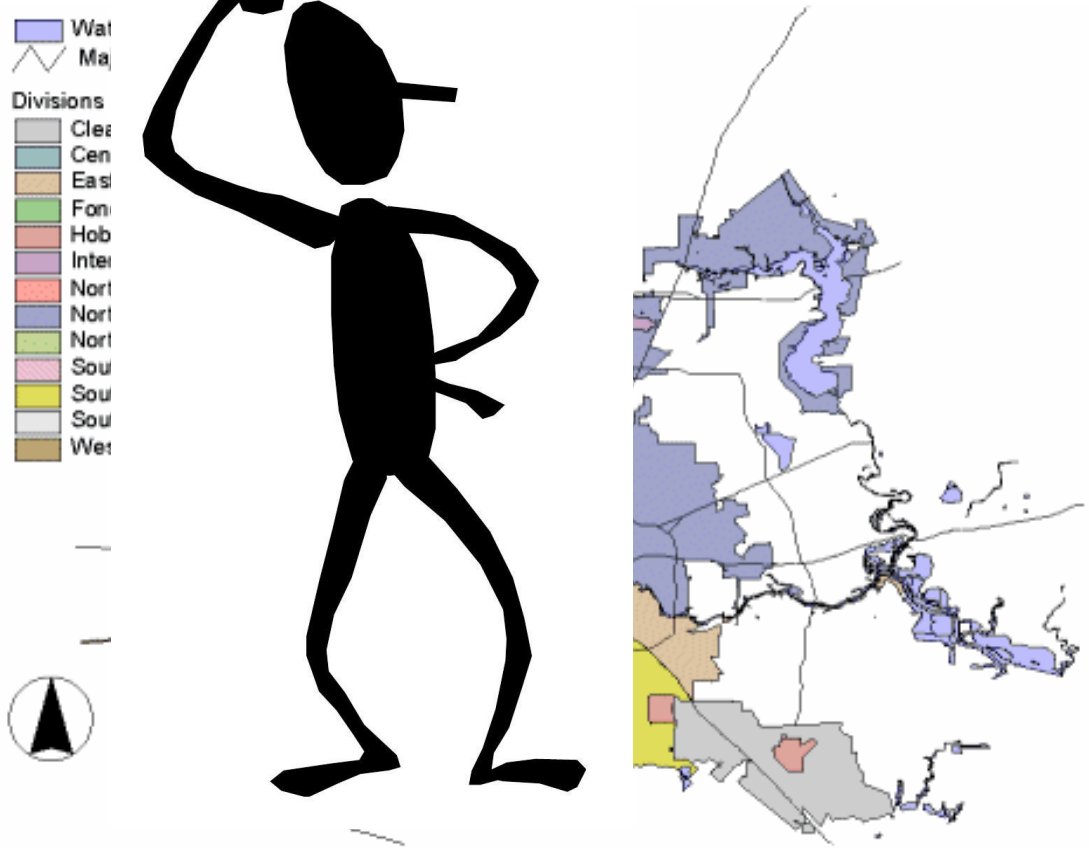
Results

- .csv with just under a million criminal offenses in Houston from June 2009 – February 2015
  812,154
- Crimes divided into city blocks with standard format for columns / block names
  (e.g., 4100-4199 FANNIN)
- Compatible with the City of Houston's enterprise GIS ROADS.shp
- ...and if anyone was curious: my district (South Central) was responsible for **5.125%** of total crimes, despite only housing **4%** of the population and covering 18.4 square miles



Results

- .csv with just under a million criminal offenses in Houston from June 2009 – February 2015 *812,154*
- Crimes divided into city blocks with standard format for columns / block names (e.g., 4100-4199 FANNIN)
- Compatible with the City of Houston's enterprise GIS ROADS.shp
- ...and if anyone was curious: my district (South Central) was responsible for 5.125% of total crimes, despite only housing 4% of the population and covering 18.4 square miles



Next Steps

Incorporating Houston 311 Data
http://hfdapp.houstontx.gov/311/index.php

- Noise complaints
- Garbage in yards
- Water / sewage outages

Putting everything online, sans ArcGIS:

- Query crime rates for given blocks from local police departments
- Assess time of day that the person wants to travel
- Route them around potentially dangerous areas

Thank you!

@nodDFW – coworking space in Dallas, TX
Open Houston: Houston's Open Data Initiative
ESCI 452 classmates

Thank you!

Questions?