* **Introduction (50 seconds | 4 Slides)**
  + We started this project asking ourselves “Is open data really open?”
    - Open data should mean that data is open and available. While data is available is it open to the public?
      * Most of the public has no idea what a comma-separated values (.csv), JavaScript Object Notation (JSON), Application Programming Interface (API) mean, let alone how to access and manipulate the files just to look at information let alone start to form understanding.
    - So, we could say that open data is not, really open at all
    - Well, we couldn’t solve that problem in the time we had and maybe as a collective we can solve it at a later date, so, we narrowed down to what we could solve, or at least help.
  + **Motivation (75 seconds | 5 Slides |** **Running Total 125 Seconds | 9 Slides)**
    - You might have noticed the title of presentation a “Geovisual Framework with National Incident Based Reporting / Open Data for Novice Users”
    - This steams from a two-phase motivation
      * First Open Data
        + Open data is data that is freely available for use and redistribution by any individual without copyright restrictions . Government agencies around the world are releasing government data in order to promote transparency, civic engagement, research and new services that benefit communities
      * National Incident Based Reporting System (NIBRS)
        + The purpose to improve the quality of crime data collected by law enforcement on each single criminal incident. The NIBRS.

Different from Universal Crime Reporting (UCR) in that it reports on each criminal incident.

* + - * + Currently 30% of all Law Enforcement Agencies (LEA)

By 2021 all LEAs are to be on the NBIRS standard

* + What this project sets out to do **(15 seconds | 1 Slide | Running Total 140 seconds | 10 Slides)**
    - Provide a framework to this:
      * Currently, no standards exist for geovisual framework to display criminal activity on digital maps for public use in an open data community.
  + How do we do this **(45 Seconds | 3 Slides | Running Total 185 Seconds | 13 Slides)**
    - Data Source
      * Our data source is the city of Dallas Open Data Portal
        + Contained within the public safety section is a dataset called “Police Incident Data”

This data contains around 400K records

103 Columns

Mainly between the years of 2015 to today.

* + - * + Our mapping focusses on burglaries

Burglary gave us an ability to segment

Residential

Commercial (Business, Government, Health etc)

Not an overwhelming number like vehicle break-ins, or a low number of activity like violent crimes.

* + - Mapping **(45 Seconds | 3 Slides | 230 Seconds | 16 Slides)**
      * Mapping is handle through open source software provided by R
        + R is a language and environment for statistical computing and graphics

We found R to be very flexible with Geographic Information System like chores.

* + - * + The actual mapping happens through a R package called Leaflet

Leaflet is the leading open-source JavaScript library for mobile-friendly interactive maps.

Leaflet is used by USA Today, The Financial Times, National Parks Service and others

* + - * + With the combination of R and Leaflet were able to create interactive publishable web maps in as little as 16 lines of code.
  + Introduction of Framework **(55 seconds | 4 Slides | Running Total 285 Seconds | 20 Slides)**
    - When we think about our audience for this framework it’s not cities like San Francisco, Dallas, Chicago, Los Angeles and New York
      * These cities are likely to have robust open data platforms that are fairly usable and kept by well-trained analysts.
      * Their officer counts are in the thousands
      * The average LEA office employs around 68 employees (<https://www.bjs.gov/content/pub/pdf/lpd13ppp.pdf>)
      * 31% are civilians that run some of the administrative functions of these LEA’s, that’s around 20 people at the average size LEA in the US.
        + This includes 911 operators, dispatchers, forensics, mechanics, HR, finance, etc., notice we did not include analysts in this listing.
  + A Geovisual Framework with NIBRS/Open Data for Open Users **(30 seconds | 2 Slides | Running Total 315 Seconds | 22 Slides)**
    - Our geovisual framework consists of 5 parts, in no particular order maps are to be
      * Time bound
      * Interactive
      * Contain a variable(s)
      * Base Map
      * Colored data
    - We found some open mapping websites operated by both public sector and private sector entities to be noisy (examples on PowerPoint), unable to get a get grasp on what is happening
  + Time **(45 Seconds | 3 slides | Running Total 350 Seconds | 25 slides)**
    - The first part of our framework is Time in a spatial context is often referred to as temporal
      * Geovisualization covering NIBRS data should be time bound
        + Victims, offenders, guardians and managers of property adjust their densities over time and around specific places, this allows from the opportunity of crime to shift and coagulate.
        + Maps should represent

A year

Month

Week

Day

* + - * + Maps can be used for comparison with different time comparisons

Think year over year, month over month and go further granular from there

* + Interactivity **(45 Seconds | 3 Slides | Running Total 395 Seconds | 28 Slides)**
    - The second part of our framework is interactivity of the map
    - Manipulation of known data is not what we are thinking about as data scientist, but the user’s private relationship with the interactivity of the map.
      * Cartographic interaction is the communication between a human and a map mediated through a computing device
      * This allows the user to ask questions, think of this like EDA as a data scientist. In the field of cartography this is known as Exploratory Spatial Data Analysis or (ESDA)
        + The user is able to zoom in and out click on a data point in some maps and gain additional areas of knowledge by interaction.
  + Color (**45 Seconds | 3 Slides | Running Total 440 Seconds | 31 Slides**)
  + Basemaps **(45 Seconds | 3 Slides | Running Total 485 Seconds | 34 Slides)**
  + Variables **(45 Seconds | 3 Slides | Running Total 530 Seconds | 37 Slides)**
  + Deployment and Conclusion **(55 Seconds | 4 Slides | Running Total 585 | 41 Slides)**