## Ideas

May be compare two states (Indiana and other state)

## Given datasets

### Questions

* How should we treat null data entries?
  + **Wait until we figure out what w b and h mean**
    - Most likely means “White, Black, and Hispanic” according to Peter.
  + For now, since some have 0s and others have nulls, that means nulls != 0s

population for female from 2005 to 2006 didn't increase by much and the number of death decreased by 7. It seems dramatic, but it's not a dramatic dip in fact.

### Kinds of prescription opioid

### 

### 

### Frequency dataset

#### Findings

**This is an incomplete dataset:**

* **While some counties have data measured every year from 1999 - 2014, there are others that are missing years of data (ex: check county 18125)**

This is fine, we just need to make sure we explain this in our report / presentation

* Showing the average over the years 1999 - 2010 gives us a better idea of which counties are struggling with drug use problems than if we just looked at individual years

Other things to mention:

* Maybe it's easier to help drug users where they are more highly concentrated
  + I'd advise using the average drug users per county vis (without factoring in population) as it's more likely to show highly concentrated areas of drug use
  + (helping by city)
* Or, maybe the Indiana government is looking for ways to divide funding among counties fairly to help decrease drug abuse. If this is the case then drug use per county population would be best at showing which counties need the most help
  + (helping by county)

47k rows

14 variables:

1. Year (1999 - 2014)
2. County
3. Drug ?
   1. w, b, h seem to never occur unless Drug does, could we just focus on the Drug values?
   2. IDEA: parse through and only analyze using values that include the 'drug' or 'opioid' and 'heroin' and omit all entries without a value in these columns
   3. **Good assumption / observation:** drugw + drugb + drugh = drug value in most cases
      1. In some cases the sum is a few values below drug though, and if w b or h are null is the sum of the non null values still equal to drug? Does this hold for opioids as well?
      2. This shows all of the county FIPS Codes from all the 52 states in the USA
         1. [https://www.nrcs.usda.gov/wps/portal/nrcs/detail/?cid=nrcs143\_0](https://www.nrcs.usda.gov/wps/portal/nrcs/detail/?cid=nrcs143_013697)

Each county has a different initial year that some data or data at all is collected.

* 1. **Amazing assumption:** w b and h might stand for white, black, hispanic
     1. Maybe the missing difference in some of the entries accounts for the 'other' category?

1. Opioid (same question as for Drug)
2. Heroin (same question as for Drug)
3. County codes for Indiana (page 136 - 156)

| 18001 | Adams | IN |
| --- | --- | --- |
| 18003 | Allen | IN |
| 18005 | Bartholomew | IN |
| 18007 | Benton | IN |
| 18009 | Blackford | IN |
| 18011 | Boone | IN |
| 18013 | Brown | IN |
| 18015 | Carroll | IN |
| 18017 | Cass | IN |
| 18019 | Clark | IN |
| 18021 | Clay | IN |
| 18023 | Clinton | IN |
| 18025 | Crawford | IN |
| 18027 | Daviess | IN |
| 18029 | Dearborn | IN |
| 18031 | Decatur | IN |
| 18033 | De Kalb | IN |
| 18035 | Delaware | IN |
| 18037 | Dubois | IN |
| 18039 | Elkhart | IN |
| 18041 | Fayette | IN |
| 18043 | Floyd | IN |
| 18045 | Fountain | IN |
| 18047 | Franklin | IN |
| 18049 | Fulton | IN |
| 18051 | Gibson | IN |
| 18053 | Grant | IN |
| 18055 | Greene | IN |
| 18057 | Hamilton | IN |
| 18059 | Hancock | IN |
| 18061 | Harrison | IN |
| 18063 | Hendricks | IN |
| 18065 | Henry | IN |
| 18067 | Howard | IN |
| 18069 | Huntington | IN |
| 18071 | Jackson | IN |
| 18073 | Jasper | IN |
| 18075 | Jay | IN |
| 18077 | Jefferson | IN |
| 18079 | Jennings | IN |
| 18081 | Johnson | IN |
| 18083 | Knox | IN |
| 18085 | Kosciusko | IN |
| 18087 | La Grange | IN |
| 18089 | Lake | IN |
| 18091 | La Porte | IN |
| 18093 | Lawrence | IN |
| 18095 | Madison | IN |
| 18097 | Marion | IN |
| 18099 | Marshall | IN |
| 18101 | Martin | IN |
| 18103 | Miami | IN |
| 18105 | Monroe | IN |
| 18107 | Montgomery | IN |
| 18109 | Morgan | IN |
| 18111 | Newton | IN |
| 18113 | Noble | IN |
| 18115 | Ohio | IN |
| 18117 | Orange | IN |
| 18119 | Owen | IN |
| 18121 | Parke | IN |
| 18123 | Perry | IN |
| 18125 | Pike | IN |
| 18127 | Porter | IN |
| 18129 | Posey | IN |
| 18131 | Pulaski | IN |
| 18133 | Putnam | IN |
| 18135 | Randolph | IN |
| 18137 | Ripley | IN |
| 18139 | Rush | IN |
| 18141 | St Joseph | IN |
| 18143 | Scott | IN |
| 18145 | Shelby | IN |
| 18147 | Spencer | IN |
| 18149 | Starke | IN |
| 18151 | Steuben | IN |
| 18153 | Sullivan | IN |
| 18155 | Switzerland | IN |
| 18157 | Tippecanoe | IN |
| 18159 | Tipton | IN |
| 18161 | Union | IN |
| 18163 | Vanderburgh | IN |
| 18165 | Vermillion | IN |
| 18167 | Vigo | IN |
| 18169 | Wabash | IN |
| 18171 | Warren | IN |
| 18173 | Warrick | IN |
| 18175 | Washington | IN |
| 18177 | Wayne | IN |
| 18179 | Wells | IN |
| 18181 | White | IN |
| 18183 | Whitley | IN |

Filtered data (Indiana Only)

<https://docs.google.com/spreadsheets/d/1iJQGNJQiE6P4Hy8y8PRXeYq49l5eyGXT/edit#gid=1667047033>

### Demographic dataset

~800 rows

6 variables:

1. State FIPS
   1. FIPS == Federal Information Processing System (Indiana's FIPS number is 18)
   2. Not useful to us, as all rows have 18 as the FIPS
2. Zip3
   1. A 3 digit ZIP code is the first three digits of a 5 digit ZIP code
   2. Using them in tableau: <https://www.reddit.com/r/tableau/comments/bdkend/comment/el0nlr8/?utm_source=share&utm_medium=web2x&context=3>
3. Year (2006 - 2016)
4. Quarter
5. Drug 9143
   1. What does this value actually represent?
      1. Is it the number of uses?
      2. **Might be the average number of uses per quarter** 
         1. And then just round up to nearest whole number
      3. That would make sense, but then why does it have decimal places?
   2. Oxycodone, is a narcotic, other names: OxyContin, Percocet, Endocet, Roxicodone, Roxicet
   3. Source: <https://www.dea.gov/sites/default/files/2020-04/Drugs%20of%20Abuse%202020-Web%20Version-508%20compliant-4-24-20_0.pdf>
6. Drug 9652
   1. Oxymorphone, is a narcotic, other names: Numorphan
   2. Source: same as for oxycodone

### MME Equivalent dataset

MME == Morphine milligram equivalent

* Doesn't have any actual data, just shows how to convert amounts of drugs ingested into equivalent amounts of morphine ingested
  + Might be useful if we want to measure the average morphine intake per county?

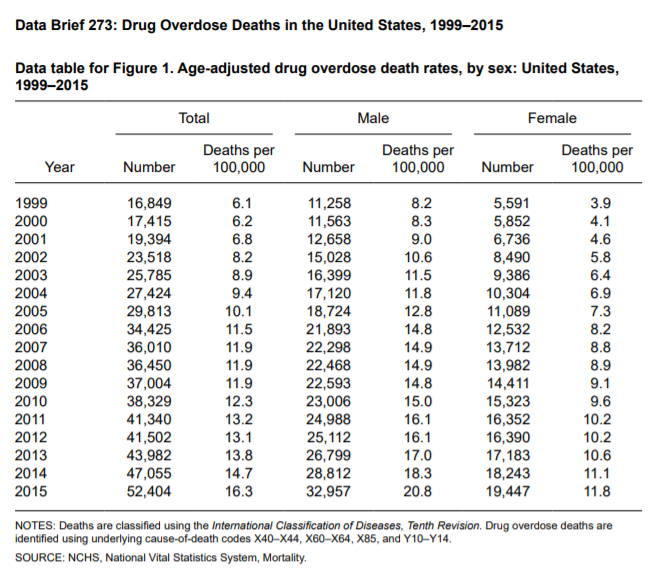
Visualization

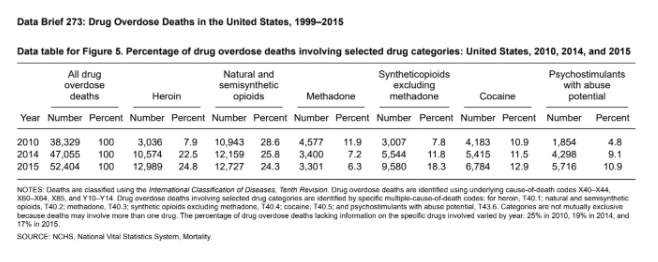
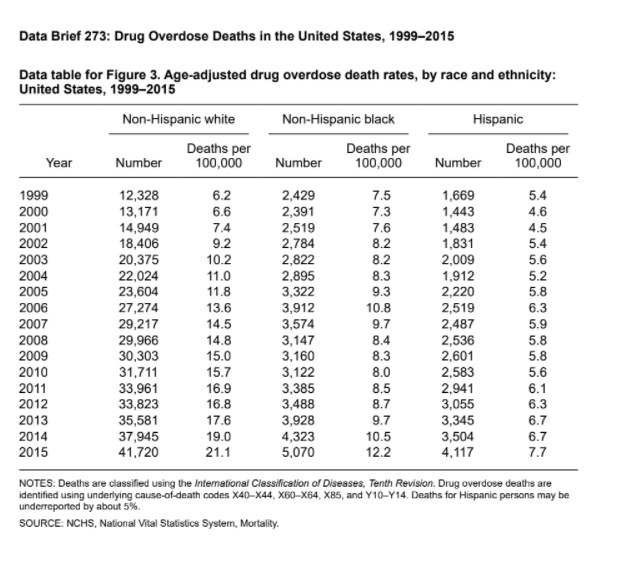
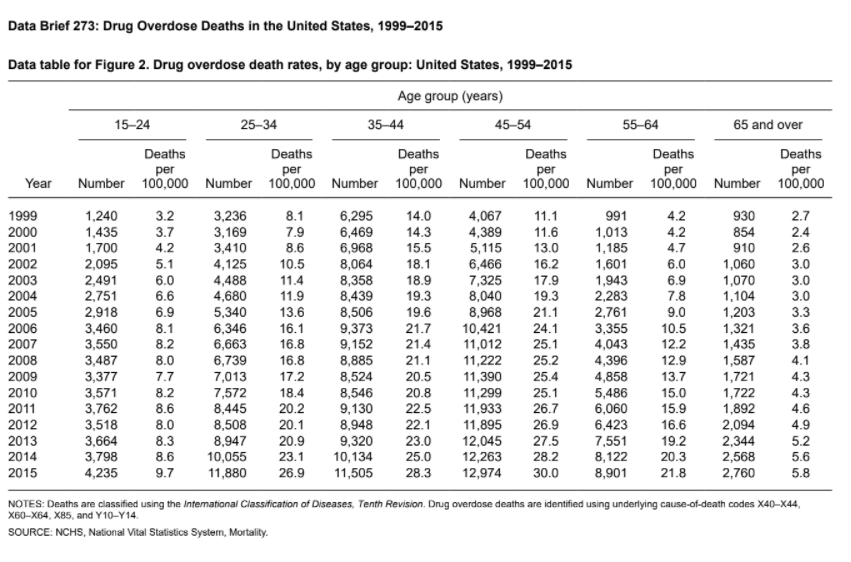
* Maps?
  + With a pie chart just like the example from the tableau trainings
* Pie Chart
  + We have some good datasets that share “race” as the common dependent variable.

## Supporting datasets

From Peter

### Examining Fatal Opioid Overdoses in Marion County, Indiana: <https://link.springer.com/content/pdf/10.1007/s11524-016-0113-2.pdf>





According to the dataset that I filtered, six counties have the most sustenance in drug, drug w, opioid, opioid w, opioid h, heroin, and heroin w. They are Allen, Lake, Delaware, Madison, Marion, and Vanderbugh. <https://1drv.ms/x/s!AkOjLxIFtOr-hnpUjwhRzU2fAj5X>

* Which drug type should we focus on? the one that has the highest sum? Or the drug that is most effective in damaging the body

### Indiana population datasets link

<https://www.stats.indiana.edu/population/> is the main website / where I found the following:

[Estimates\_90s.xls](https://www.stats.indiana.edu/population/estimates_90s.xls) has 1990 - 1999 by county for Indiana

[Intercensal00s.xlsx](https://www.stats.indiana.edu/population/intercensal00s.xlsx) has 2000 - 2010 by county for Indiana

What I did to the data:

* Removed the april 1st 2000 and april 1st 2010 population entries (the actual censuses), as it wasn't necessary (have jul 1st 2000 - 2010 *estimates*)
* Only used the jul 1st 1999 entry from the 1990 - 1999 dataset since the given frequency dataset is from 1999 - 2014
* Excluded years 2011 - 2014 from the population visualization since that data has not been released yet in aggregate