1. Predict which areas do not currently have smoke alarms
   1. Measure 1: model (Greg)
      * ***target* *variable*:** % of home Red Cross Visits that led to a smoke alarm install
        + 80%/20% dichotomous again or continuous 0-100%?
        + data source: Red Cross plus Census (ACS 2009-2013)
        + block groups or census tracts?
          1. if census tracts, need to modify Brian’s code to change from block group focus
      * ***predictors***:
        + county level census data for training
        + tract/zip level census data for generating predictions
   2. Measure 2: model (Greg)
      * ***target variable:*** % of fire incidents within a region that didn’t have smoke detectors (Xianghui create variable, push to Greg for model)
      * ***predictors***:
        + county level census data for training
        + tract/zip level census data for generating predictions
        + ACS data will be available at Google Drive\Red Cross\Data\Census\acs.csv or can be downloaded from enigma github (Find acs.csv)
      * data source: NFIRS fireincidents & basicincidents
      * challenge: presence of smoke detectors is missing or unknown “U” for many cases.
      * Possible solutions: hierarchical modeling; multiple imputation; within states proportional the % of fire departments known to report data.
   3. Measure 3: Indicator: (Andrew)
      * ***indicator:***% of homes in area that don’t have working smoke alarm
      * data source: AHS survey
      * self reported data. known to underestimate true effect.
      * We only have AHS smoke detector at the county level for a small subset of counties (some entire states are missing). Unless we adopt Brian’s methodology of building a model to predict this indicator for the missings, I don’t think we have a value add using the indicator as-is. I’d advocate simply picking up Brian’s predictions.
2. Which areas are in most danger of having fires
   1. Measure 1: indicator & time series model (Xianghui)
      * ***indicator*:** # of incidents per 1000 people within a region
      * challenge: how to deal with regions where fire departments don’t report data
      * data source: NFIRS
   2. Measure 2: indicator & time series model (Xianghui)
      * ***indicator***: # of incidents Red Cross responds to per 1000 people within a region
      * data source: Red Cross disaster dataset
      * challenge: Red Cross responses are not perfectly representative of where fires happen.
      * Time series modeling predicting incidents in next period based on historical and seasonality
   3. Measure 3: indicator (Andrew)
      * ***indicator:*** # of RC fire responses per region
      * data source: RC disaster data

Note these constructed could be reported at a relatively high level (county). Additionally, we could build models to predict these outcomes at the county level and then score more granular regions (tracts), if we have the granular level data. Identifying, quantifying and acquiring data on the drivers of fire incidents could also be tough.

Can look at NFIRS data over the past 5 years for trends by census tract

Potentially add in season

1. Where are people most at risk of injury and death (David Marx)
   1. Measure 1: Indicator & model
      * ***indicator:***# of injuries per 1000 people
      * data source: NFIRS fire
      * ***predictors***:
        + county level census data for training
        + tract/zip level ACS census data for generating predictions
   2. Measure 2: Indicator & model
      * ***indicator:***# of deaths per 1000 people
      * data source: NFIRS fire
      * ***predictors***:
        + county level census data for training
        + tract/zip level ACS census data for generating predictions

Note these constructed could be reported at a relatively high level (county). Additionally, we could build models to predict these outcomes at the county level and then score more granular regions (tracts), if we have the granular level data. Identifying, quantifying and acquiring data on the drivers of injuries and deaths related to fire incidents could also be tough.

----------------------------------------------------------------------------------------------------------------------

Open questions for 09/21/2015 worknight:

* Is it feasible to build a model predicting fire incidents/deaths/injuries at county level and use it to score census tracts?
* what are the drivers of fire incidents/deaths/injuries?
* is this data available from census at tract level?
* How does the report which the states the % of fire departments data per state calculate this figure? Can we use the underlying data or somehow figure out which regions have fire departments that are not reporting data to NFIRS?
* Are there any census variables that are predictive of fires below the tract level?
* Unit of analysis-previously at the county level (took averages at the county level-for example, average age of a county)
* <https://github.com/enigma-io/smoke-alarm-risk>

Hackpad-<https://hackpad.com/American-Red-Cross-DJuXtxOtqvw>

Predictors at the census tract level-research-KC

Next call on October 5th at 7PM to check in and determine next steps, recruiting new volunteers

Other potential variables of interest:

-privately rented home (more likely to not have a smoke alarm in the home) vs. owner occupied-

<http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC1067731&blobtype=pdf>

-Homes with at least one high school graduate were nearly four times more likely to be fully

protected. Homes that had multiple levels, a basement, or were cluttered or poorly cleaned were

significantly less likely to be fully protected.

<http://www.researchgate.net/publication/7442095_When_one_is_not_enough_prevalence_and_characteristics_of_homes_not_adequately_protected_by_smoke_alarms>

having a central heating source

<http://link.springer.com/article/10.1023%2FA%3A1010478116532>