**Source Datasets**

* Ventyx retail sales data (2011, 2010, and 2009)
  + These were the three most up to date versions available at the time of analysis
  + Data includes total annual retail sales of electricity at the electric service territory level (by state) for the following sectors: residential, commercial, industrial, transportation, all
  + Specific attributes of interest for each sector include total annual sales (in MWH), total annual revenue (in thousands of $), and total customers
* Ventyx Electric Service Territory Boundaries (2011)
  + These were intersected with Ventyx State and Province Boundaries to split territories across state boundaries. This makes the polygons consistent with the reporting level of the Ventyx retail sales data
* EIA 2011 State Level data, Tables 1-4 (<http://www.eia.gov/electricity/sales_revenue_price/>)
  + Table 1: Number of consumers (bundled and unbundled) by sector, Census Division, and State
  + Table 2: Sales to bundled and unbundled consumers by sector, Census Division, and State
  + Table 3: Revenues for sales to bundled and unbundled consumers (including delivery service revenue) by sector, Census Division, and State
  + Table 4: Average retail price for bundled and unbundled consumers by sector, Census Division, and State
  + The data in these tables corresponds to the data in the Ventyx retail sales data; however, they are at a different spatial resolution (state level data, not electric service territories)
* Landscan USA Raster Datasets (2011 pre-release)
  + These are from the HSIP 2012 dataset and are the most recent version available. The vintage also corresponds roughly to the vintage of the Ventyx and EIA data
  + Include (nominal) 90 m resolution rasters for the continental US, AK, and HI showing daytime and nighttime population

**Data Preparation: Backfilling Ventyx Retail Sales Data**

* The Ventyx 2011 retail sales data represents the most current and fine spatial resolution estimates of annual electric load by broad market sectors available for the DG Wind study. As such, our goal was to use this as the primary data source, and backfill any data gaps from other appropriate datasets.
* First, we joined the Ventyx 2011 retail sales data for industrial, commercial, and residential sectors (including attributes for # of consumers, sales in MWH, and revenue in thousands of $) to the Ventyx 2011 Electric Service Territory Boundaries. Upon review, we identified three types of data gaps:
  + Areas not covered by any electric service territories:
    - Upon review, many of these were in areas with very little population and/or in areas designated as wilderness or parks (e.g., National Parks, National Forests).
    - We therefore assumed these were valid gaps where very little load was likely to be present, and ignored them
  + Areas with Electric service territories with reported load of zero:
    - Due to overlaps in electric service territories, there were very few areas where this occurred for residential. The only major population center with this problem was San Francisco, for which the only applicable service territory was the City and County of SF utility. Upon further review, we determined that PGE also served residential customers in SF, and therefore simply adjusted the PGE territory boundary to cover northern SF.
    - Need to check this for commercial and industial?
  + Areas with Electric Service Territories but no reported load from the Ventyx 2011 Retail Sales data
    - This was a fairly common problem (>700 territories of various sizes, but due to overlaps, actual gaps were fewer) and therefore treated it as a reporting omission. The remaining discussion explains how such gaps were backfilled
* Where data was not reported in the 2011 Ventyx retail sales data, we backfilled using the following data source (in order): 2010 Ventyx retail sales data, 2009 Ventyx retail sales data, 2011 EIA state level data
  + 11 service territories were backfilled using the 2010 Ventyx data
  + 10 service territories were backfilled using the 2009 Ventyx data
  + 712 service territories were backfilled using the 2011 EIA state level data
* For residential load, the effects of backfilling were fairly minimal compared to the total US annual load:
  + The 2010 Ventyx data backfill accounted for 12125024 mwh, or 0.83% of the total US load
  + The 2009 Ventyx data backfill accounted for 11603129 mwh, or 0.8% of the total US load
  + The 2011 EIA state data backfill accounted for 49479810 mwh, or 3.4% of the total US load
* For the 2011 EIA state level data backfill, we first aggregated the 2011 Ventyx retail load by state, then subtracted it from the EIA state total. We did not include the 2009 and 2010 Ventyx loads in the state-level aggregation because we found that the nationwide totals of 2011 Ventyx retail load and 2011 EIA state level retail load were very similar (difference of 7 mwh, or 0.0000005%). In some cases, subtracting the Ventyx aggregated load from the EIA state totals resulted in negative load. This represents a strange data discrepancy into which we have very little insight. As such, we simply ignored negative load values in any subsequent analyses.

**Residential Load Disaggregation**

* In order to calculate residential load at county resolution, we first disaggregated residential load from the electric service territory level to a nominal 90m resolution grid for the US. We assumed that residential locations were represented by the 2011 Landscan night time population rasters, and assumed that each person used the same amount of electricity.
* Prior to performing the disaggregation, we performed an analysis to refine the 2011 Landscan night time population rasters, eliminating areas that had nonzero night time population, but were unlikely to be residential. Examples of such areas include prisons, hotels/motels, college campuses, and commercial and industrial facilities that operate 24 hours a day. To mask these areas out, we:
  + Calculated a grid showing the ratio of nighttime to daytime population using the Landscan 2011 daytime and nighttime population grids.
  + Reclassified the ratios into two groups: nighttime/daytime <= 1 🡪nonresidential, and nighttime/daytime > 1 🡪 residential.
    - Results are stored in: S:\mgleason\DG\_Wind\Data\Analysis\residential\_load\residential\_land\_masks\nightpop\_to\_daypop\_ratio
  + Merged the following files from HSIP/NAVTEQ 2012: Prison Areas (Polygons), College Boundaries (Polygons), Industrial Complexes (Polygons), and Hotels-Motels (Points) buffered by 90m (in postgres)
    - Results are stored in: S:\mgleason\DG\_Wind\Data\Analysis\residential\_load\residential\_land\_masks\non\_res\_facilities
  + Converted the merged polygons to raster (in Postgres and Arc), where presence of a polyon had a value of 1.
    - Results are stored in: S:\mgleason\DG\_Wind\Data\Analysis\residential\_load\residential\_land\_masks\non\_res\_facilities
  + Performed map algebra to eliminate both the is-res ratio rasters and the non-res facilities rasters from nighttime pop:
    - Con(("non\_res\_us.tif" == 1) | ("us\_isres" == 0),0,"conus\_night")
  + This had the effect that all areas classified as nonresidential were reassigned a “night time population” to zero.
  + Due to intricacies in the load disaggregation process that came later, this raster was then clipped to the outer boundaries of wind\_ds.county\_geom (performed in ArcGIS) to ensure that all residential locations were within the bounds of the county geometries
    - 🡪
  + Using the same data, we also created a simple “residential land mask” (where 1 = residential, and NoData = not):
    - Con(("us\_isres" == 1) & ("non\_res\_us" == 0),1)
    - 🡪F:\data\mgleason\DG\_Wind\Data\Analysis\residential\_load\residential\_land\_masks\combined\_mask\res\_mask\_us
  + For conus, the output of the previous step was resampled to a 0.0018 degree (~200 m) resolution grid in WGS84 consistent F:\data\mgleason\aws\_2014\_update\gis\conus\cf\_rasters\derived\onshoremask

(using nearest neighbor resampling)

* + - 🡪 F:\data\mgleason\DG\_Wind\Data\Analysis\residential\_load\residential\_land\_masks\revised\_2014\_02\_05\res\_mask
  + The output from this was the clipped using Extract By Mask to the onshoremask raster:
    - F:\data\mgleason\DG\_Wind\Data\Analysis\residential\_load\residential\_land\_masks\revised\_2014\_02\_05\res\_mask\_clip
    - \*\*\* This is the final residential land mask raster
  + The output from the previous step was then converted to individual points, which were then exported to a CSV, and these points were loaded to postgres as likely residential locations:
    - F:\data\mgleason\DG\_Wind\Data\Analysis\residential\_load\residential\_land\_masks\revised\_2014\_02\_05\res\_mask.gdb\res\_mask\_pts
    - F:\data\mgleason\DG\_Wind\Data\Analysis\residential\_load\residential\_land\_masks\revised\_2014\_02\_05\res\_mask\_pts.csv
  + These rasters were then loaded into Postgres as the “residential population rasters” for the subsequent steps:
* To perform the disaggregation, we first found the total residential population within each electric service territory. For each electric service territory, we then used the total population to calculate the proportion of people located at each grid cell within the territory. Next, we allocated the total residential load for the territory to the grid cells based on the proportion of people located at each cell. Finally, to deal with overlapping service territories, we mosaicked the grids for each service territory together, summing up the load for any overlapping cells.
* **Notes/Areas for improvement:** 
  + For CONUS, 3,489,388mwh of load were lost during disaggregation from the service territories. This load was based on backfilling sliver gaps around state borders with EIA state totals. The biggest ones are around CA (919598.000000000000 mwh) and around southern MD/northern WV (1810007 mwh).