**DESCRIPTION**

Predict cancer mortality rates for US counties.

**SUMMARY**

**Multiple Linear Regression Challenge**

**Background**

These data were aggregated from a number of sources including the American Community Survey ([census.gov](http://census.gov/)), [clinicaltrials.gov](http://clinicaltrials.gov/), and [cancer.gov](http://cancer.gov/). Most of the data preparation process can be veiwed [here](https://data.world/nrippner/cancer-trials).

**Your Task: Build a multivariate Ordinary Least Squares regression model to predict *"TARGET\_deathRate"***

**Deliverables**

a. Your final model equation

b. The statistical software output including (adjusted) R-squared and Root Mean Squared Error (RMSE)

c. Your code file (if you used a programming language)

d. Model diagnostics including statistics and visualizations:

- Assess linearity of model (parameters)

- Assess serial independence of errors

- Assess heteroskedasticity

- Assess normality of residual distribution

- Assess multicollinearity

e. Your interpretation of the model

f. Other factors to consider:

- Are there any outliers?

- Are there missing values?

- How will you handle categorical variables?

**Good Luck and Have Fun!**

**Data Dictionary**

**TARGET\_deathRate:** Dependent variable. Mean *per capita* (100,000) cancer mortalities(*a*)

**avgAnnCount:** Mean number of reported cases of cancer diagnosed annually(*a*)

**avgDeathsPerYear:** Mean number of reported mortalities due to cancer(*a*)

**incidenceRate:** Mean *per capita* (100,000) cancer diagoses(*a*)

**medianIncome:** Median income per county (*b*)

**popEst2015:** Population of county (*b*)

**povertyPercent:** Percent of populace in poverty (*b*)

**studyPerCap:** *Per capita* number of cancer-related clinical trials per county (*a*)

**binnedInc:** Median income per capita binned by decile (*b*)

**MedianAge:** Median age of county residents (*b*)

**MedianAgeMale:** Median age of male county residents (*b*)

**MedianAgeFemale:** Median age of female county residents (*b*)

**Geography:** County name (*b*)

**AvgHouseholdSize:** Mean household size of county (*b*)

**PercentMarried:** Percent of county residents who are married (*b*)

**PctNoHS18\_24:** Percent of county residents ages 18-24 highest education attained: less than high school (*b*)

**PctHS18\_24:** Percent of county residents ages 18-24 highest education attained: high school diploma (*b*)

**PctSomeCol18\_24:** Percent of county residents ages 18-24 highest education attained: some college (*b*)

**PctBachDeg18\_24:** Percent of county residents ages 18-24 highest education attained: bachelor's degree (*b*)

**PctHS25\_Over:** Percent of county residents ages 25 and over highest education attained: high school diploma (*b*)

**PctBachDeg25\_Over:** Percent of county residents ages 25 and over highest education attained: bachelor's degree (*b*)

**PctEmployed16\_Over:** Percent of county residents ages 16 and over employed (*b*)

**PctUnemployed16\_Over:** Percent of county residents ages 16 and over unemployed (*b*)

**PctPrivateCoverage:** Percent of county residents with private health coverage (*b*)

**PctPrivateCoverageAlone:** Percent of county residents with private health coverage alone (no public assistance) (*b*)

**PctEmpPrivCoverage:** Percent of county residents with employee-provided private health coverage (*b*)

**PctPublicCoverage:** Percent of county residents with government-provided health coverage (*b*)

**PctPubliceCoverageAlone:** Percent of county residents with government-provided health coverage alone (*b*)

**PctWhite:** Percent of county residents who identify as White (*b*)

**PctBlack:** Percent of county residents who identify as Black (*b*)

**PctAsian:** Percent of county residents who identify as Asian (*b*)

**PctOtherRace:** Percent of county residents who identify in a category which is not White, Black, or Asian (*b*)

**PctMarriedHouseholds:** Percent of married households (*b*)

**BirthRate:** Number of live births relative to number of women in county (*b*)

(*a*): years 2010-2016

(*b*): 2013 Census Estimates