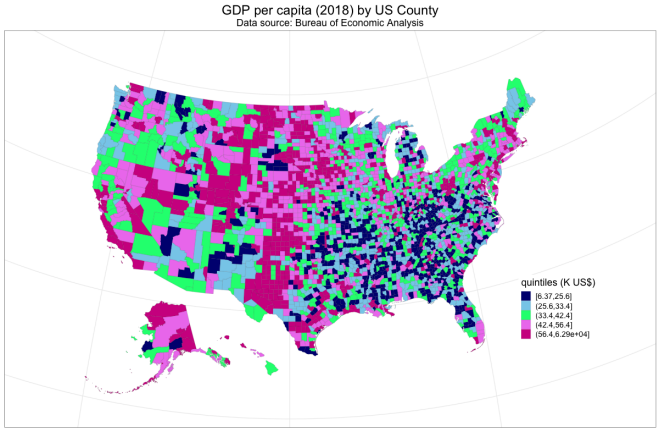
Analysis of GDP Data for US Counties Using Machine Learning

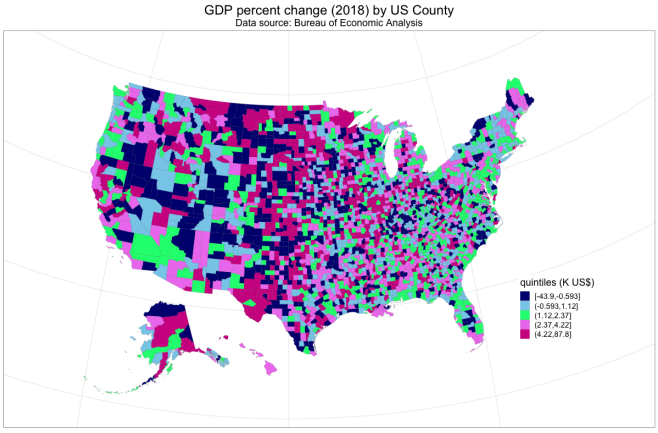
July 31, 2020
K Durham

Introduction

As we find ourselves in the year 2020 dealing with unprecedented challenges, it has been observed that the United States population is experiencing an ever-increasing fissure between economic extremes. In his theory for how we got here, Morgan Housel from The Collaborative Fund [1] puts forth a concise yet comprehensive economic history of the United States since WW2, in which he outlines some of the drivers for our current material disparities. While in political terms it has become common to refer to America as either 'red' or 'blue' at the state level, John B. Judis (Washington Post Magazine) [2], in analyzing results from the 2018 midterm elections, notes the appearance of widening 'blue dots' neighboring US cities surrounded by otherwise red background regions. He suggests that economic trends, particularly those impacting employment in America's suburbs, may offer an explanation.

In an attempt to explore potential drivers for and to give some geographical clarity to these inequalities, this report employs a systematic approach to examine the newly-available Real Gross Domestic Product (Real GDP) data by US county released by the Bureau of Economic Analysis in December 2019. From a simple examination of these data overlaid on a map of the US, it is apparent that summarizing GDP at the state level, as with recent political patterns, may over-smooth informative variation for understanding American economic life. Unlike metropolitan areas, which are city-based, US counties represent comprehensive, continuous areas covering both urban and rural parts of all 50 states and the District of Columbia. While some heterogeneity at the county level surely still exists, this level of segmentation allows for diversity to be examined at a finer scale than when using data summarized at the state level. The analyses presented in this report are motivated by a dual objective: 1) To explore potential drivers for the differences in GDP that are observed at the county level, and 2) To use these variables to predict future GDP levels and GDP change by county for the US.





Summary

Objectives and Data The objectives of this analysis were to identify potential drivers of US county-level GDP and to explore the possibility of using these variables to predict future year GDP. The GDP data from BEA were divided by population census data for each county to obtain per capita estimates. A limitation of this analysis was the amount of relevant data available for all counties to explain GDP. Here we use: 1) broad indicators of the economic condition of a region, including information on employment, poverty and construction, 2) population and migration information, 3) data reporting the number of establishments and annual payroll per county by industry sector, and 4) US Census regions and divisions.

Descriptive Summary To live in a county with the highest per capita GDP, move to Texas! Among a ranking of counties by GDP per capita, the top 10 are almost exclusively dominated by counties in this state (which is likely driven in part by low population numbers in its western region). The ranking also reveals that the bottom 10 counties are consistently from the US Southeast, predominantly the states of Kentucky and Mississippi. Solace for the Southeast may be that counties from this region also appear frequently among those not just with the lowest GDP but also with the highest quintile of growth rates. In general, the total GDP per capita by county top 10 are relatively stable year over year for this data period. In contrast, the GDP percent change per capita rankings are much more dynamic. Interestingly, for percent change, both the top 10 and bottom 10 tables are dominated by counties from the same states located in the Western and Central US, including North and South Dakota, Nebraska, Montana and Texas, perhaps indicating increased volatility in this region.

Modeling Results Using data from 2016 - 2018, county GDP per capita for the following year was predicted with high precision (R-squared = 99%) when using current year GDP-derived variables in the model. When current year GDP was removed, the R-squared reduced to 76%. Other variables that were important in explaining variation in per capita GDP were annual payroll and the number of establishments per capita, the proportion of the number of establishments and payroll allocated to retail trade, and the proportion of the number of establishments allocated to wholesale trade. None of the data sets or models used in this analysis were able to predict future year percent change in GDP per capita with reasonable precision.

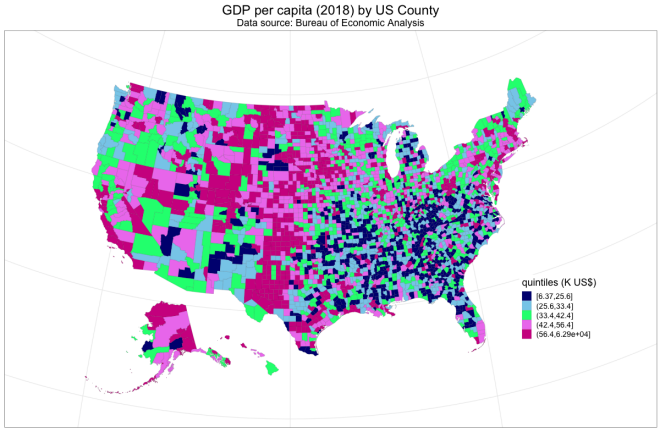
GDP Associated Variables The County Business Patterns data were transformed so that the total number of establishments and the annual payroll would represent per capita values. Also, the values by sector were converted to percentages of the county total values, hopefully revealing any sectors whose relative domination of the business environment for a county led to higher or lower GDP per capita. County GDP increases with the total pay and number of establishments, as well as with the percentage of a county's establishments allocated to wholesale trade. While this includes the outputs of 'agriculture, mining, manufacturing, and certain information industries, such as publishing,' this sector may be associated with the recent growth in e-commerce, where 'warehousing' has been transformed to the final step before the consumer receives a product directly. On the flip side, county GDP appears to decrease the larger the percentage of its establishments and payroll are dedicated to retail, and to a smaller extent, manufacturing.

Data

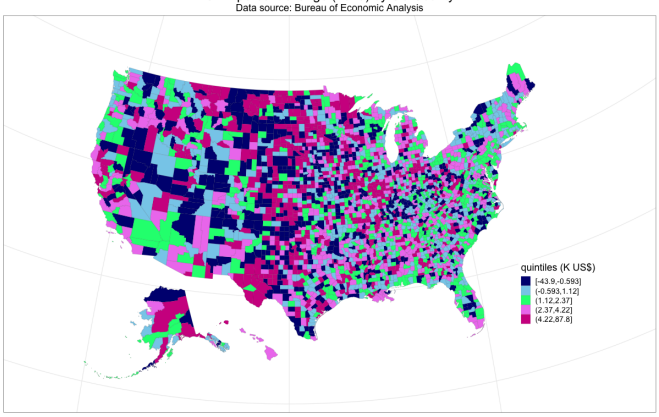
Data were collected for this report with objectives to predict and explore GDP county economic patterns and their prime movers. While the US government provides a wealth of data, it can be challenging to obtain all-inclusive information at the county level. The explanatory variables used in these analyses were primarily of 4 classes 1) broad indicators of the economic condition of a region, including information on employment, poverty and construction, 2) population and migration information, including the net international migration to a county, 3) data from County Business Patterns (CBP), which reports the number of establishments and annual payroll per county by industry sector, and 4) US Census regions and divisions, to allow for larger geographic trends to explain GDP county patterns. Data were obtained primarily from the Bureau of Economic Analysis (BEA) [3] (GDP) and the US Census Bureau (USCB) [4], with additional data on employment patterns obtained from the US Bureau of Labor Statistics (USBLS) [5]. GDP data were available for 3113 counties for the years 2016 - 2018, thus data for other variables were also assembled for this time frame. The variables in tables 9 and 10 were available for virtually all counties, and were combined with GDP data to form the data set for analysis. CBP data was transformed so that the values for number of establishments and annual pay by sector were converted to their percentage of the total for all sectors for a county. The total number of establishments and total annual pay were then divided by the county population, resulting in per capita values. One caveat for data analysis concerns GDP data for the state of Virginia, where only aggregate data were available for certain geographic areas. In these cases, the total GDP amount was divided evenly by the number of distinct areas included, which certainly results in sub optimal estimates. However, it does ensure that the total GDP for the region remains consistent. Figure 1 displays choropleth plots of the US by county for both GDP per capita and percent change for the years 2016 - 2018.

Even solely from a descriptive standpoint, these data sets provide rich opportunities for exploration. For example, tables 1 and 2 display the top 10 and bottom 10 counties as ranked by GDP per capita, respectively, for the years 2016 - 2018. Similar tables for the percent change in GDP (tables 11 and 12) are contained in the Appendix. While the tables are sorted by the indicated measure, both GDP per capita and percent change are included on all tables for comparison. In general, the per capita totals tables top 10 are relatively stable year over year for the data period. Interestingly, for each year studied the county with the highest per capita GDP is Loving, Texas, the second least populated county in the US (after Kalawao, Hawaii) with a 2018 population of 152, resulting in a per capita GDP estimate an order of magnitude higher than the next highest value. For this reason, it was the only county excluded from modeling. The relatively low population of West Texas is also the likely driver behind that region's dominance of the top GDP per capita table. On the other side of the rankings, the lowest per capita GDP values were found in the Southeast region of the US, predominantly Kentucky and Mississippi, with appearances by Tennessee, West Virginia and Georgia. However, the silver lining for this region is that it is also the home of counties in the bottom quintile in terms of GDP per capita that also exhibited multiple years in the top quintile of GDP growth for the period. These counties may represent interesting opportunities to examine growth drivers and differentiating factors from neighboring regions. Table 3 displays these counties, located predominately in the US South and West, along with the number of years they met these criteria and their GDP statistics for the last year of the trend.

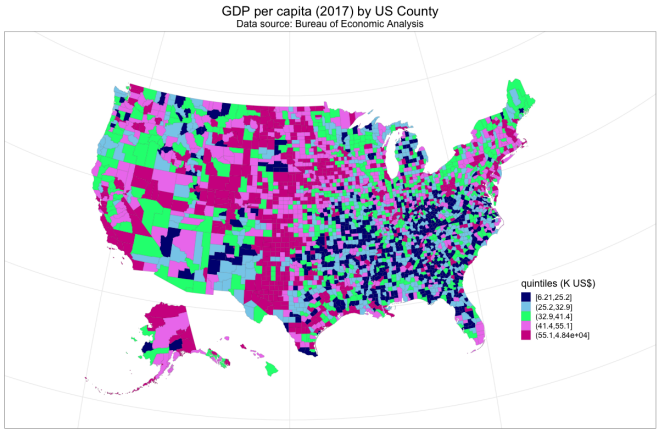
In contrast to the total GDP per capita values per county, the GDP percent change per capita rankings are much more dynamic from year to year. Interestingly, both the top 10 and bottom 10 tables are dominated by counties from the same states located in the central US, including North and South Dakota and Nebraska, indicating an interesting pattern of economic variability in the region.



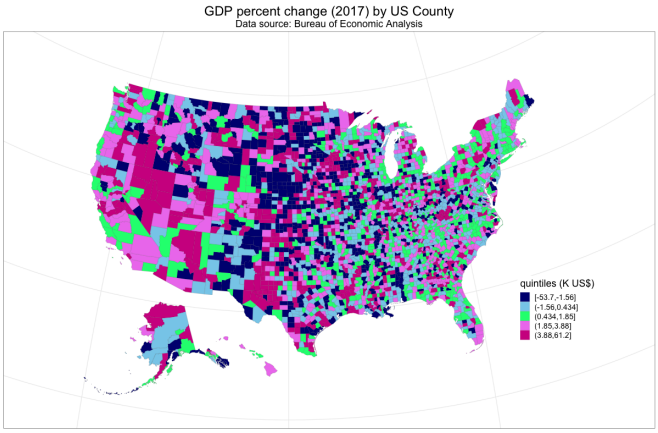
(c) GDP per capita, 2018



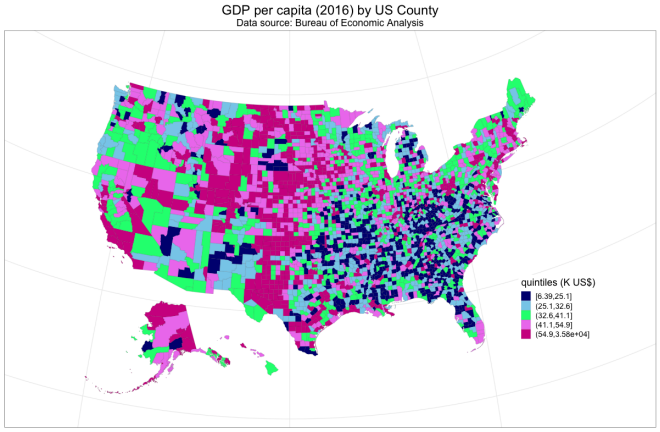
(d) GDP percent change, 2018



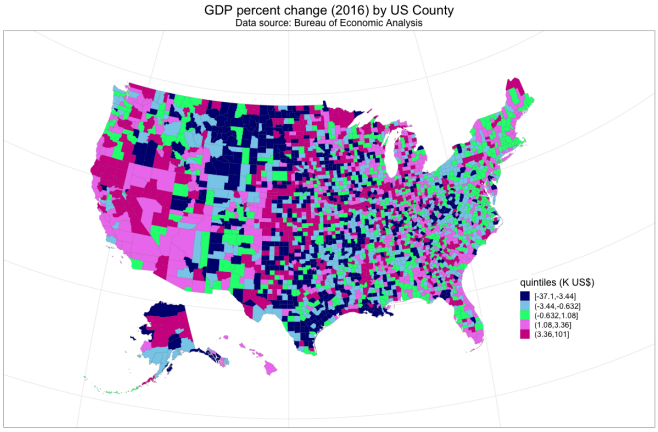
(e) GDP per capita, 2017



(f) GDP percent change, 2017



(g) GDP per capita, 2016



(h) GDP percent change, 2016

Figure 1: GDP per capita and percent change by year

year	GDP per capita	GDP percent change	County and State
2016	35799.74	78.38	Loving Texas
2016	5211.26	-5.83	McMullen Texas
2016	2509.04	7.22	Glasscock Texas
2016	1487.86	12.92	Upton Texas
2016	1429.95	-17.08	Roberts Texas
2016	1166.49	-15.46	Irion Texas
2016	1131.71	19.18	Reagan Texas
2016	1087.39	-4.31	North Slope Alaska
2016	960.04	21.31	Martin Texas
2016	884.35	-19.95	La Salle Texas
2017	48428.71	35.28	Loving Texas
2017	4922.41	-5.54	McMullen Texas
2017	3080.13	22.76	Glasscock Texas
2017	1847.12	24.15	Upton Texas
2017	1433.43	26.66	Reagan Texas
2017	1307.35	-8.57	Roberts Texas
2017	1259.26	31.17	Martin Texas
2017	1221.63	4.73	Irion Texas
2017	1142.42	5.06	North Slope Alaska
2017	978.34	10.63	La Salle Texas
2018	62876.55	29.83	Loving Texas
2018	4981.15	1.19	McMullen Texas
2018	3923.17	27.37	Glasscock Texas
2018	2007.51	8.68	Upton Texas
2018	1703.2	35.25	Martin Texas
2018	1597.82	11.47	Reagan Texas
2018	1128.35	-7.64	Irion Texas
2018	1124.73	-13.97	Roberts Texas
2018	1060.53	-7.17	North Slope Alaska
2018	1034.27	5.72	La Salle Texas

Table 1: Top 10 Counties by GDP per capita and year

Modeling and Analysis

To perform modeling to meet the analysis objects, training and validation data were developed according to the following:

1. The data were divided into sets of explanatory variables corresponding to those measured in 2016 and 2017.

2. GDP variables were assigned as the response for the following year, i.e. independent 2016 variables were combined with GDP from 2017 as the dependent variable, and similarly for 2017 independent

year	GDP per capita	GDP percent change	County and State
2016	11.61	-5.77	Garrard Kentucky
2016	11.55	-4.95	Morgan Tennessee
2016	11.31	-1.36	Nicholas Kentucky
2016	11.24	1.46	Carroll Mississippi
2016	11.12	-7.44	Magoffin Kentucky
2016	10.84	-3.95	Greene Mississippi
2016	10.65	3.92	Spencer Kentucky
2016	10.18	-1.14	Wirt West Virginia
2016	9.41	1.58	Elliott Kentucky
2016	6.39	-1.82	Long Georgia
2017	11.59	-0.26	Owsley Kentucky
2017	11.44	-4.86	Jackson Kentucky
2017	11.4	-1.83	Garrard Kentucky
2017	11.32	4.44	Greene Mississippi
2017	10.85	-4.01	Nicholas Kentucky
2017	10.73	-3.46	Magoffin Kentucky
2017	10.56	-0.79	Spencer Kentucky
2017	9.63	2.33	Elliott Kentucky
2017	9.61	-5.64	Wirt West Virginia
2017	6.21	-2.82	Long Georgia
2018	11.59	-0.04	Owsley Kentucky
2018	11.53	0.85	Jackson Kentucky
2018	11.39	0.61	Greene Mississippi
2018	11.28	5.09	Magoffin Kentucky
2018	10.86	0.08	Nicholas Kentucky
2018	10.86	-8.89	Lincoln West Virginia
2018	10.3	-2.51	Spencer Kentucky
2018	9.97	3.79	Wirt West Virginia
2018	9.52	-1.18	Elliott Kentucky
2018	6.37	2.47	Long Georgia

Table 2: Bottom 10 Counties by GDP per capita and year

County and State	Last Year	GDP per capita	GDP percent change	No. Years
Antrim Michigan	2018	23.99	3.98	2
Assumption Louisiana	2018	20.35	4.75	2
Bates Missouri	2017	25.15	5.38	2
Bledsoe Tennessee	2018	14.26	6.68	2
Boise Idaho	2017	23	14.03	2
Bullock Alabama	2018	24.09	4.95	2
Caldwell Missouri	2018	25.05	14.94	2
Calhoun Georgia	2017	21.06	8.74	2
Cleveland Arkansas	2018	21.18	6.07	2
Cook Georgia	2018	21.37	7.65	2
Coryell Texas	2017	20.06	4.64	2
Cumberland Kentucky	2018	23.78	7.61	2
Daviess Missouri	2017	22.7	5.81	2
Dillon South Carolina	2018	24.85	4.59	2
East Feliciana Louisiana	2018	22.95	4.63	2
Elbert Colorado	2018	18.42	4.93	2
Falls Texas	2018	22.96	3.97	2
Greene North Carolina	2017	24.5	5.08	2
Henry Indiana	2018	23.5	6.86	2
Jenkins Georgia	2018	15.64	4.84	2
Leake Mississippi	2018	22.49	4.59	2
Lee South Carolina	2017	19.76	6.66	2
Lincoln Arkansas	2018	19.51	5.02	2
Lyon Kentucky	2017	24.1	8.68	2
McCormick South Carolina	2018	20.17	3.99	2
Menard Illinois	2018	23.53	15.24	2
Morgan Indiana	2018	22.11	6.66	2
Ottawa Kansas	2018	22.18	4.42	2
Perry Ohio	2018	19.87	4.53	2
Randolph Arkansas	2017	22.32	5.81	2
Sabine Texas	2017	24.82	4.37	2
Starke Indiana	2017	20.02	4.43	2
Stewart Georgia	2018	21.2	6.58	2
Talbot Georgia	2018	21.77	12.66	2
Valencia New Mexico	2018	16.5	4.66	2
Vinton Ohio	2018	21.68	9.19	2
Wasatch Utah	2017	24.04	4.7	2
Washita Oklahoma	2018	22.4	4.26	2
West Carroll Louisiana	2018	23.01	6.54	2

Table 3: Growth Counties

variables and 2018 GDP.

3. The response variables GDP per capita and GDP per capita percent change from previous year were computed for each county.

4. The 2016 independent / 2017 dependent data was designated as the training data, and the 2017 independent / 2018 dependent data was assigned as validation data for model development.

5. Data pre-processing (described below) was developed using the training data and subsequently applied to the test data.

Data Pre-processing

Before analysis the data were processed as follows:

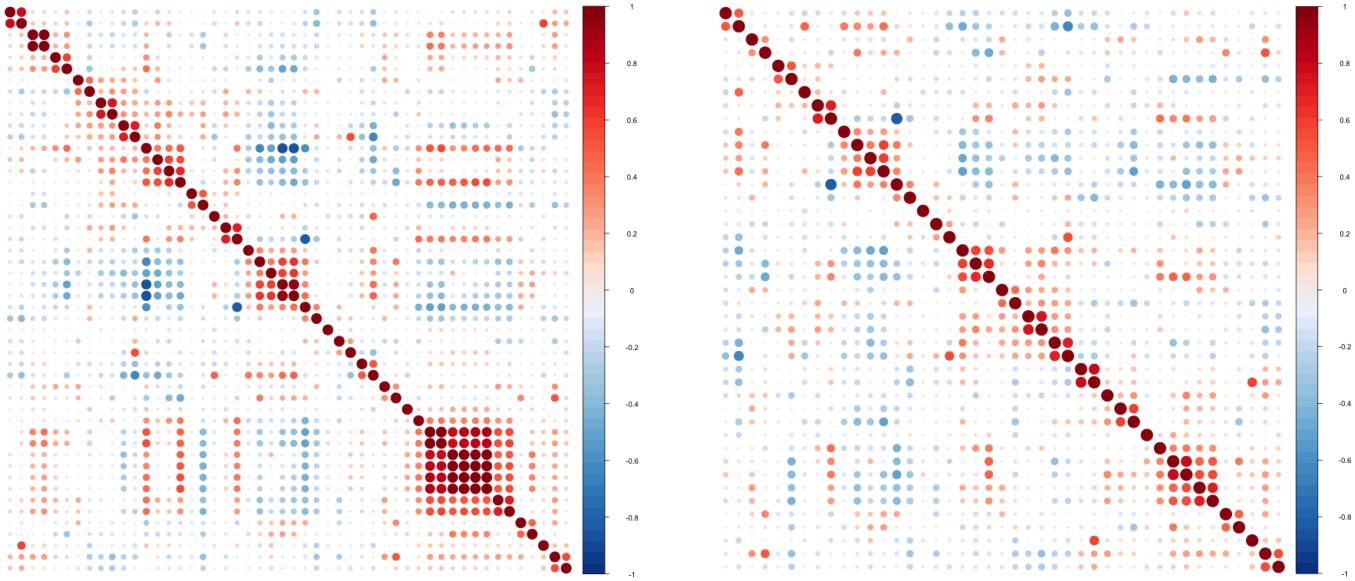
1. Variables were assessed for missing values and deleted if the percentage of counties missing was greater than a 10% threshold. A collection of variables (n=21), all from County Business Patterns was removed at the step. This was expected, as these data are often intentionally not included at the county level due to the lack of anonymity for the small sample size. The removed variables were: ann_payK_admin, ann_payK_agri, ann_payK_arts, ann_payK_edu, ann_payK_info, ann_payK_manage, ann_payK_mining, ann_payK_nc, ann_payK_re, ann_payK_utilities, no_estab_admin, no_estab_agri, no_estab_arts, no_estab_edu, no_estab_info, no_estab_manage, no_estab_mining, no_estab_nc, no_estab_re and no_estab_utilities (see table 10).

2. The correlations among the variables were computed and variables that were highly correlated with others ($r > 0.85$) were omitted from the model fits. In this analysis, 7 variables were removed: bldgs_value, med_house_inc, pov_under18, employed, labor_force, pop_est and netmig_rate (see table 9). Figure 2 displays the correlation matrices before and after removal of these variables. Once the important variables from the models were identified, the correlations of the important features with those omitted were examined.

3. Data were centered and scaled and any remaining missing values were filled in using median imputation.

4. GDP per capita as a response was log-transformed before analysis; GDP percent change was not transformed.

After all processing, the analysis data set contained 43 variables. All pre-processing and modeling was performed using the caret package in R (as described in Applied Predictive Modeling [6]).



(a) Before

(b) After

Figure 2: Correlation among Variables

Models

A collection of 3 linear models (Linear Model [LM], Partial Least Squares [PLS] and Elastic Net [ENET]) and 3 non-linear models (Support Vector Machines [SVM], Multivariate Adaptive Regression Splines [MARS] and Gradient Boosted Machines [GBM]) were fit to both the Counties Subset and All Counties data for 4 different analyses:

GDP per capita All variables for the prior year were included to predict GDP per capita for the following year, including GDP-derived variables.

GDP per capita, no prior GDP In order to examine the effects of other factors besides GDP to predict GDP per capita in the following year, GDP-related variables for the prior year were omitted for this analysis.

GDP percent change All variables for the prior year were included to predict GDP percent change for the following year.

GDP percent change compounded To deal with the apparent noise in the GDP percent change data year over year, a 3-year compounded 'return' was computed using the percent change data for the years 2016 - 2018. Variables for the year 2015 were used in the models to see if they could explain this forward-looking trend. For this model, a cross-sectional training / test set split was randomly performed, with 60% of data in the training set and 40 % in the test set.

For each model, 10-fold cross validation (CV) was performed for the training set to determine the optimal values of tuning parameters (except for the linear model, which has none). Distributions of R-squared, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) were reviewed for the training set, along with single values of these quantities available for the validation set. Observed validation set responses were plotted against predicted values, and variables with top importance scores across multiple models were also reported and visualized. The important variables were also examined for correlation with features that had been removed in the pre-processing step.

Results

Table 4 provides links to the figures and tables summarizing the results for all six models fit for different analyses where applicable. In summary, the analyses results were as follows:

GDP per capita GDP per capita was able to be predicted with high precision for 2018 using data from the previous year. For these data, model choice was critical, with the non-linear options resulting in an increase in R-squared values of between 23-30 percentage points over the linear models. GBM led the way with an R-squared value of 0.99. Because the response was analyzed as natural log of GDP per capita, the RMSE approximates the percentage error for the model, i.e. a 7% error in predicting GDP per capita for the GBM model. In addition to the previous year's GDP, other important variables included annual payroll and the number of establishments per capita and the percent of a county's payroll allocated to retail trade. None of the highly correlated variables that were removed were associated with the important variables as identified by the models.

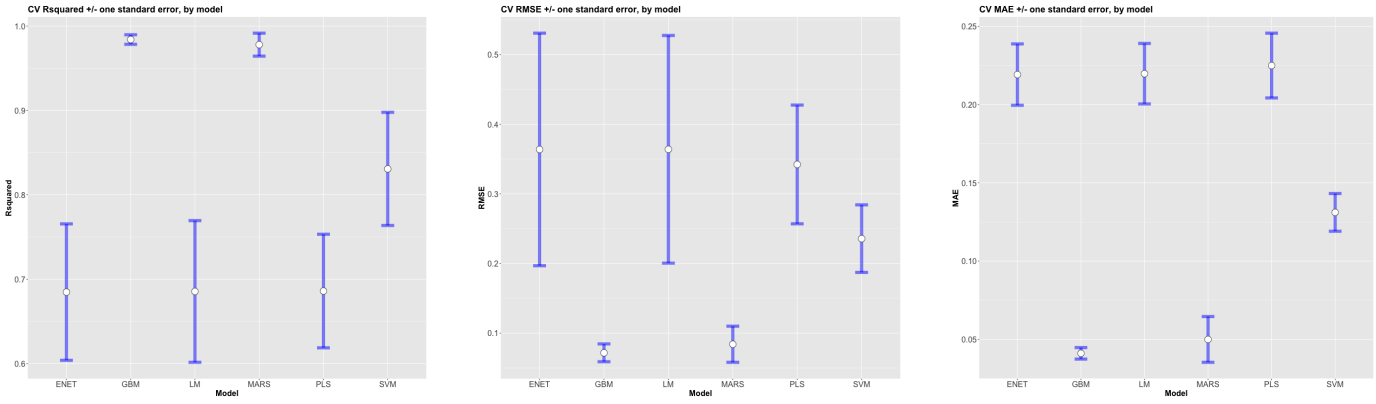
GDP per capita, no prior GDP When prior year GDP-derived variables were removed from the model, the proportion of variance explained dropped to 0.76 as computed using the SVM and GBM

models. Again, the non-linear models resulted in an approximate 20 percentage point improvement over the linear options. The important variables for these models included annual payroll and the number of establishments per capita, the proportion of the number of establishments and payroll allocated to retail trade, and the proportion of the number of establishments allocated to wholesale trade.

GDP percent change None of the data sets or models used in this analysis were able to predict future year percent change in GDP per capita with reasonable precision (data not shown). Training set values for R-squared of only 5-10% could be achieved, with test set values of about 0% for GDP percent change and 2-4 % for the composite approach.

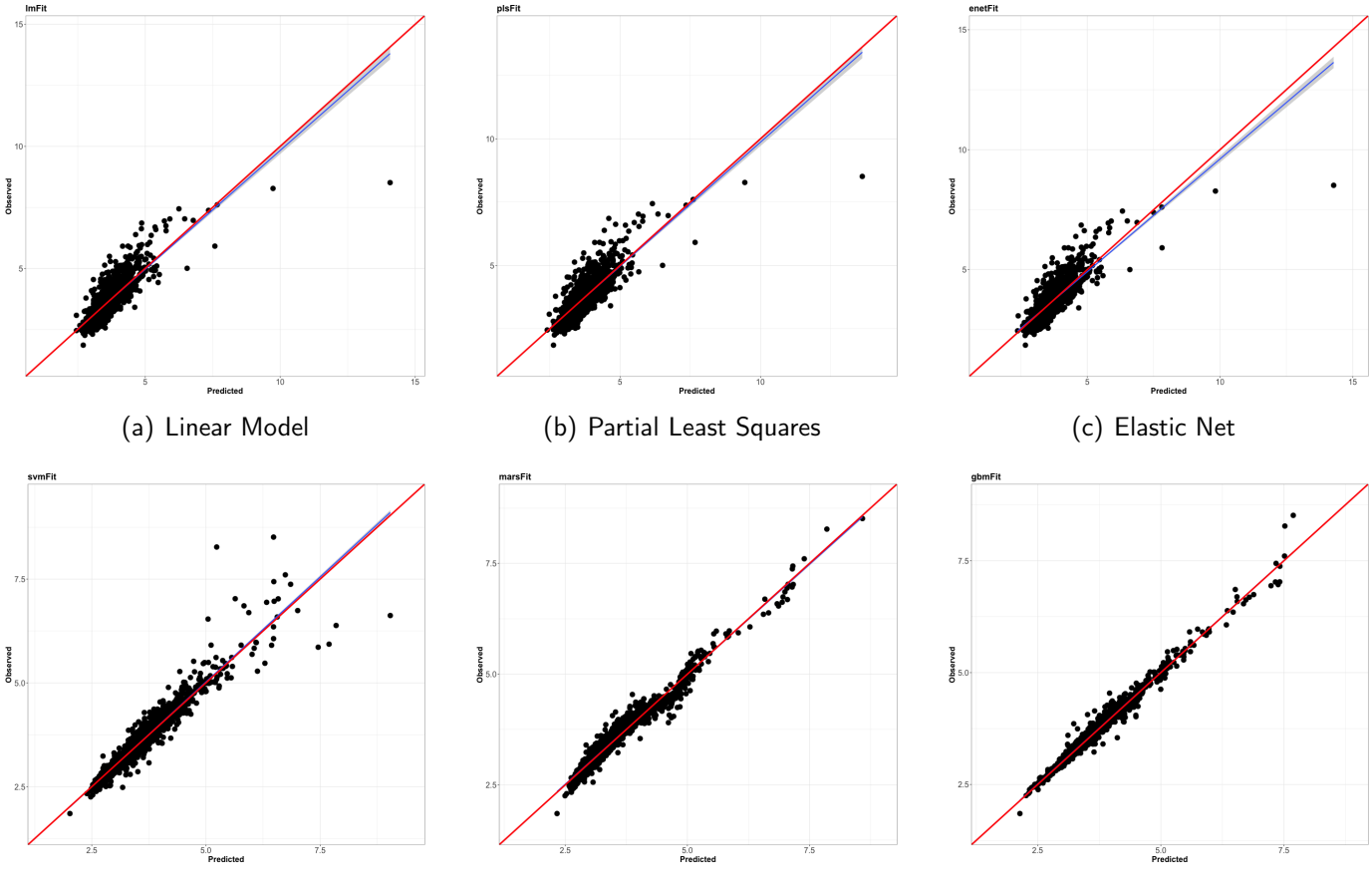
Analysis	CV	Validation Performance	Imp Var
GDP per capita	Figure 3	Figure 4 and Table 5	Figure 5 and Table 6
GDP per capita, no GDP var	Figure 6	Figure 7 and Table 7	Figure 8 and Table 8

Table 4: Results by Analysis and Data Set



(a) R-squared (b) RMSE (c) MAE

Figure 3: GDP per capita, Training Set CV Performance Measures



(a) Linear Model (b) Partial Least Squares (c) Elastic Net

(d) Support Vector Machine (e) MARS (f) Gradient Boosted Machine

Figure 4: GDP per capita, Validation Set, Observed vs. Predicted

	lm	pls	enet	svm	mars	gbm
RMSE	0.32	0.32	0.32	0.16	0.11	0.07
Rsquared	0.69	0.68	0.69	0.92	0.96	0.99
MAE	0.21	0.22	0.21	0.09	0.08	0.04

Table 5: GDP per capita, Validation Set Performance Measures

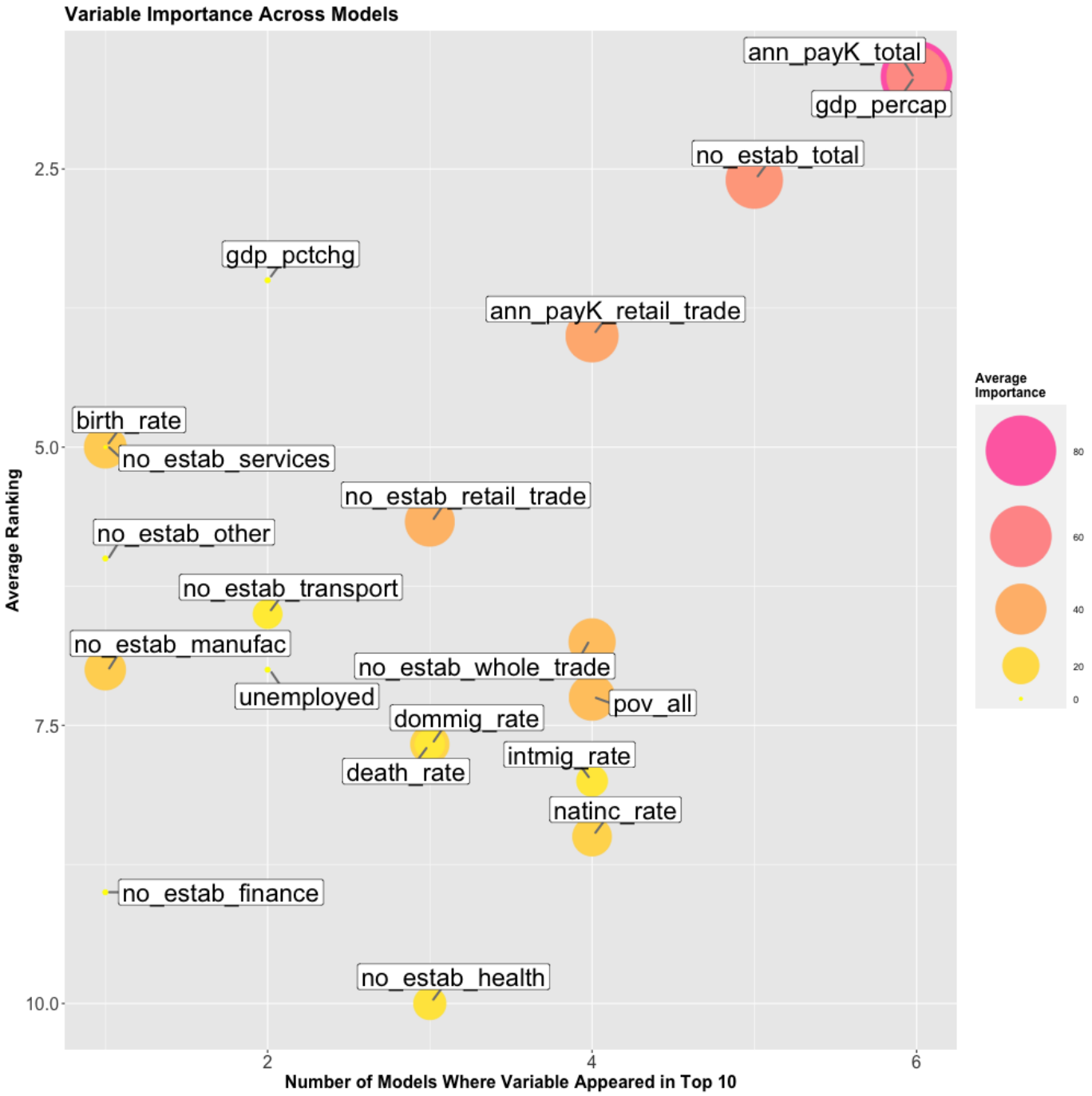
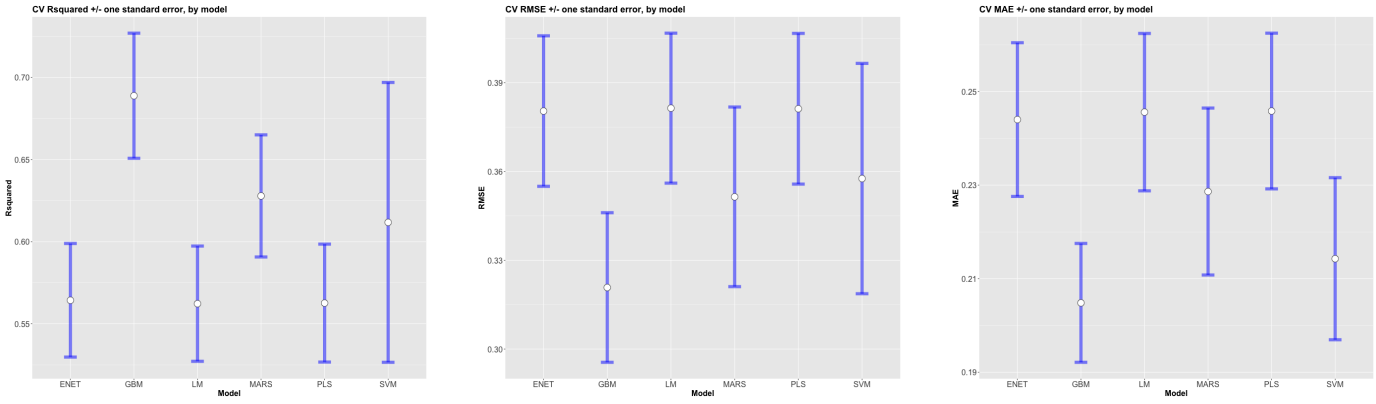


Figure 5: Variable Importance, GDP per capita

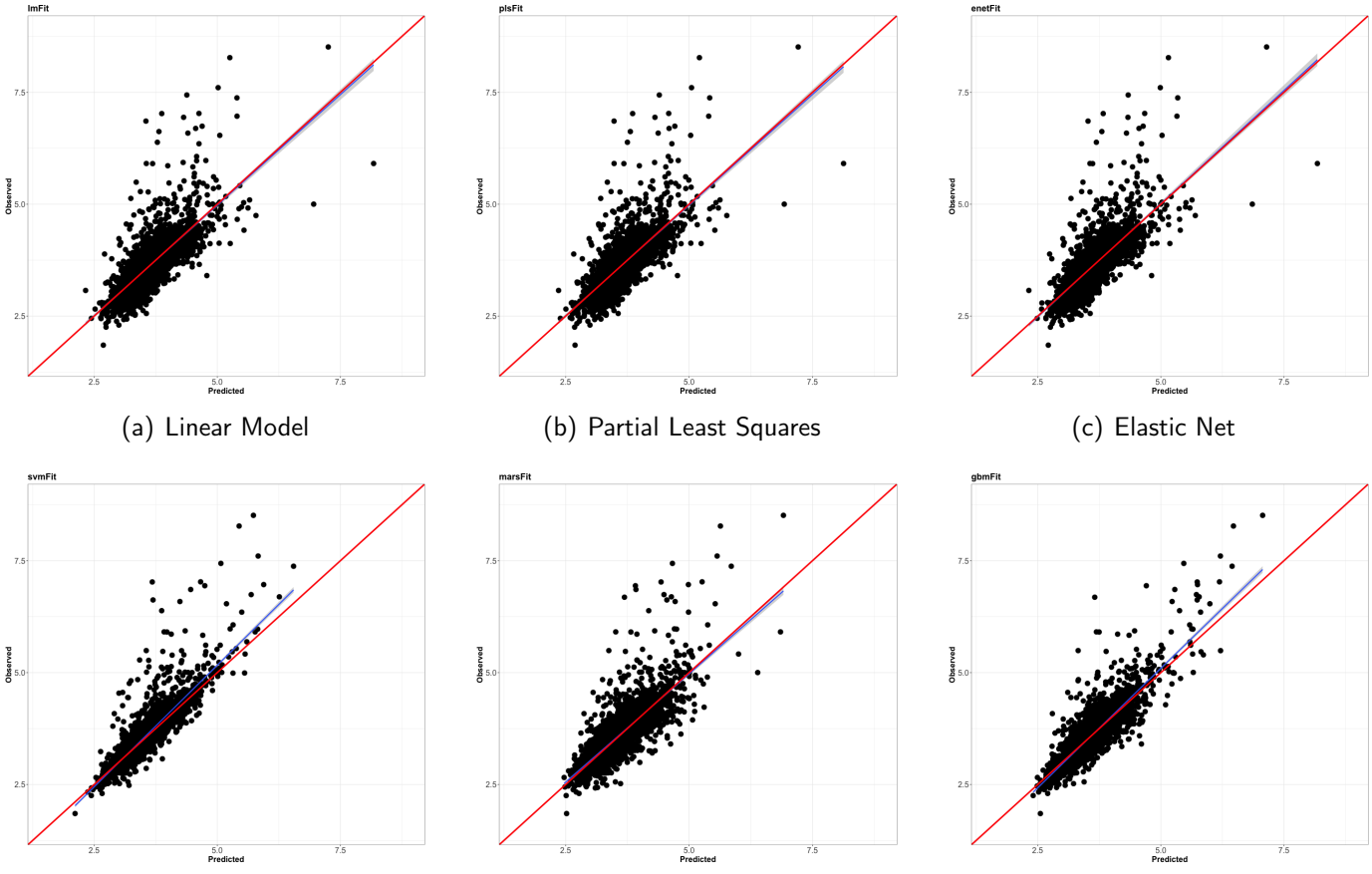
var	avg_imp	avg_rank	num_models
gdp_percap	83.02	1.67	6
ann_payK_total	57.61	1.67	6
no_estab_total	51.58	2.6	5
ann_payK_retail_trade	43.65	4	4
no_estab_retail_trade	38.11	5.67	3
no_estab_whole_trade	33.77	6.75	4
pov_all	32.77	7.25	4
no_estab_services	27.32	5	1
no_estab_manufac	25.37	7	1
natinc_rate	23.01	8.5	4
dommig_rate	22.61	7.67	3
no_estab_health	15.53	10	3
intmig_rate	14.06	8	4
death_rate	12.99	7.67	3
no_estab_transport	12.14	6.5	2
gdp_pctchg	0.08	3.5	2
no_estab_other	0.08	6	1
no_estab_finance	0.07	9	1
unemployed	0.07	7	2
birth_rate	0	5	1

Table 6: GDP per capita, Important Variables in Multiple Models



(a) R-squared (b) RMSE (c) MAE

Figure 6: GDP per capita, no GDP variables, Training Set CV Performance Measures



(a) Linear Model (b) Partial Least Squares (c) Elastic Net

(d) Support Vector Machine (e) MARS (f) Gradient Boosted Machine

Figure 7: GDP per capita, no GDP variables, Validation Set, Observed vs. Predicted

	lm	pls	enet	svm	mars	gbm
RMSE	0.37	0.37	0.37	0.29	0.35	0.29
Rsquared	0.57	0.57	0.57	0.76	0.63	0.76
MAE	0.24	0.24	0.24	0.15	0.22	0.19

Table 7: GDP per capita, no GDP variables, Validation Set Performance Measures

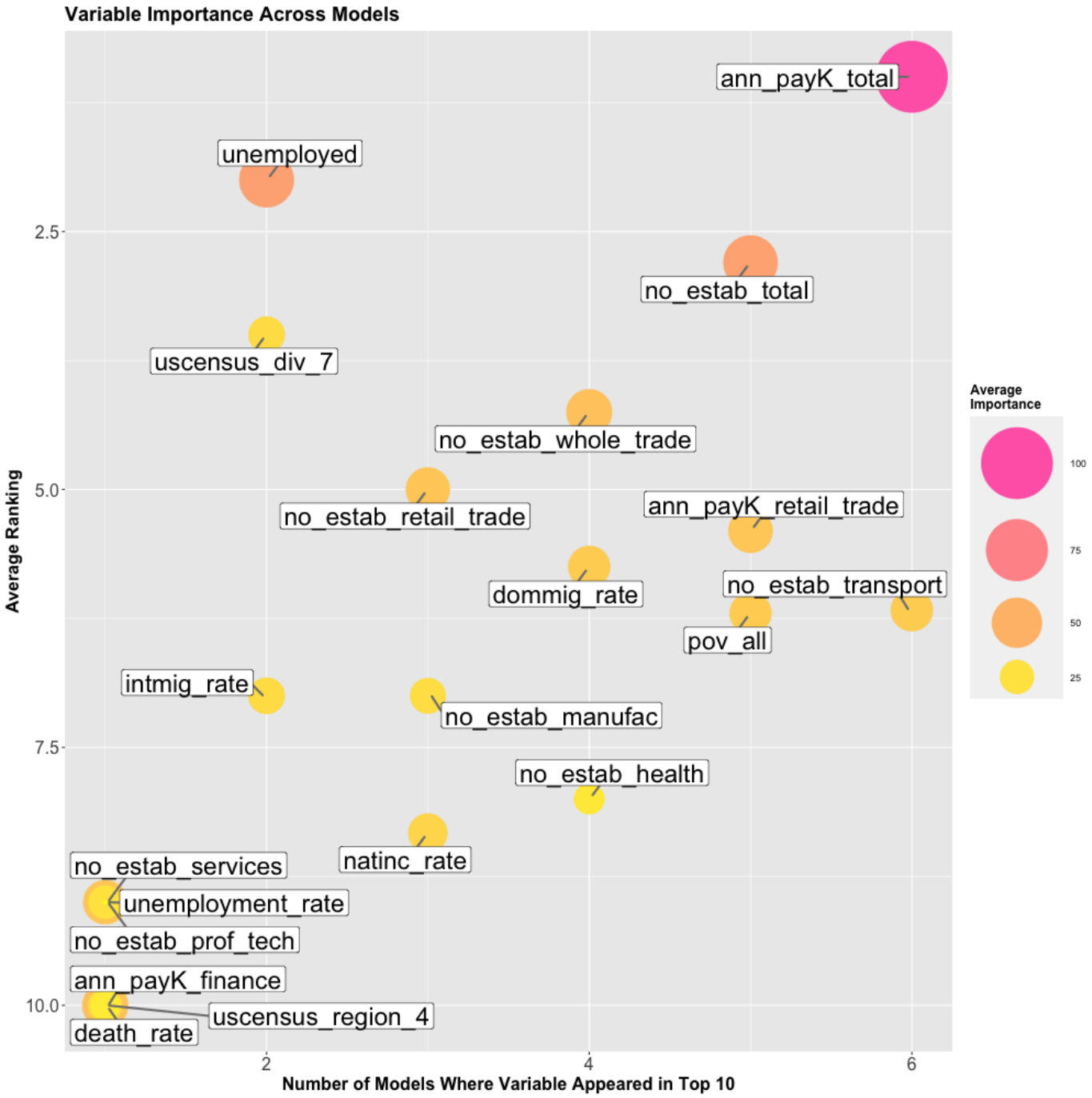


Figure 8: Variable Importance, GDP per capita, no GDP variables

var	avg_imp	avg_rank	num_models
ann_payK_total	100	1	6
unemployed	59.22	2	2
no_estab_total	58.28	2.8	5
death_rate	42.44	10	1
no_estab_whole_trade	42.2	4.25	4
ann_payK_retail_trade	39.68	5.4	5
no_estab_services	39.61	9	1
no_estab_retail_trade	39.29	5	3
no_estab_transport	36.37	6.17	6
dommig_rate	36.23	5.75	4
pov_all	35.92	6.2	5
natinc_rate	31.8	8.33	3
intmig_rate	28.09	7	2
uscensus_div_7	27.92	3.5	2
no_estab_manufac	26.96	7	3
ann_payK_finance	25.14	10	1
unemployment_rate	24.55	9	1
no_estab_health	21.02	8	4
uscensus_region_4	20.22	10	1
no_estab_prof_tech	6.23	9	1

Table 8: GDP per capita, no GDP variables, Important Variables in Multiple Models

Discussion

In this analysis, while a county's GDP per capita appeared fairly stable and relatively easy to predict, predicting how much a county's GDP would change in a subsequent year proved difficult. At first glance this seems counter-intuitive, but may have to do with the scale of the data. The GBM prediction model with an R-squared of 99% indicated an error rate of around 7% when predicting a county's GDP per capita value. While we have featured the top and bottom 10 counties out of interest, the vast majority of counties have an annual percent change that is closer to zero (from the quintiles breakdown of percent change for 2018, the middle 60% of values is in the range (-0.6, 4.2). It's also possible that looking at yearly changes for GDP is similar to looking at daily stock prices. The long-term trends definitely exist, but with only 3 year of data at this point, the noise may be too great to model the relationship meaningfully.

The County Business Patterns data was transformed so that the total number of establishments and the annual payroll would represent per capita values. Also, the values by sector were converted to percentages of the county total values, hopefully revealing any sectors whose relative domination of the business environment for a county led to higher or lower GDP per capita for that county. The variable importance information in figure 8 and table 8 for the models without GDP-derived explanatory variables indicate that annual payroll and the number of establishments per capita, the proportion of the number of establishments and payroll allocated to retail trade, and the proportion of the number of establishments allocated to wholesale trade are important, but do not give the direction of their effect. Figure 9 displays the rank correlations between log GDP per capita for 2018 and important explanatory variables for 2017. These indicate that county GDP increases with the total pay and number of establishments, as well as with the percentage of a county's establishments allocated to wholesale trade. This NAISC (North American Industry Classification System) grouping corresponds to 'establishments engaged in wholesaling merchandise, generally without transformation, and rendering services incidental to the sale of merchandise.' While this includes the outputs of 'agriculture, mining, manufacturing, and certain information industries, such as publishing,' this sector may be associated with the recent growth in e-commerce, where the 'warehousing' has actually been transformed to the final step before the consumer receives a product directly. Interesting, county GDP appears to decrease the larger the percentage of its establishments and payroll are dedicated to retail, and to a smaller extent manufacturing.

Finally, thank you for reading this report on what has proven to be a very interesting and complicated topic! Fortunately, the quality and amount of data should only increase from here, as will our efforts to understand and interpret the information. If you'd like to discuss further, please email me: katey@dahliaanalytics.com

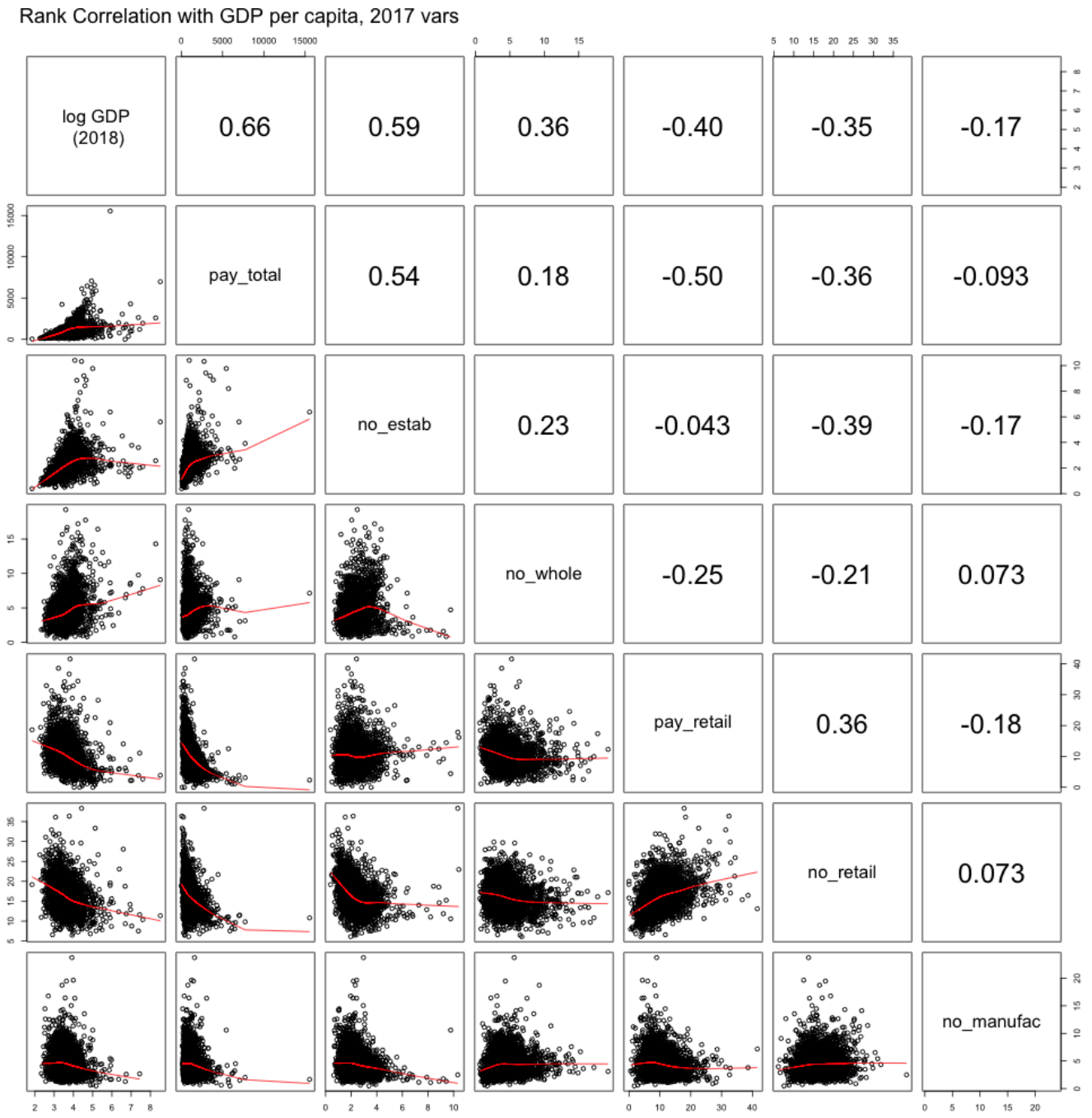


Figure 9: Correlation of GDP per capita with Important Variables

Appendix

Variable	Description	Source
division	Division	USCB
gdp	Prototype Real Gross Domestic Product, chained (2012) dollars (K)	BEA
pop_est	Resident Total Population Estimate	USCB
birth_rate	Birth Rate	USCB
death_rate	Death Rate	USCB
natinc_rate	Natural Increase Rate	USCB
intmig_rate	Net International Migration Rate	USCB
dommig_rate	Net Domestic Migration Rate	USCB
netmig_rate	Net Migration Rate	USCB
labor_force	Labor Force	USBLS
employed	Employed	USBLS
unemployed	Unemployed	USBLS
unemployment_rate	Unemployment Rate	USBLS
pov_all	Poverty Percent All Ages	USCB
pov_under18	Poverty Percent Under Age 18	USCB
med_house_inc	Median Household Income	USCB
bldgs	Building Permits	USCB
bldgs_value	Building Permits Value	USCB

Table 9: Data Variables

Annual payroll ($1,000)	Number of establishments	Description
ann_payK_admin	no_estab_admin	Admin and support and waste management
ann_payK_agri	no_estab_agri	Agriculture, forestry, fishing and hunting
ann_payK_arts	no_estab_arts	Arts, entertainment, and recreation
ann_payK_construct	no_estab_construct	Construction
ann_payK_edu	no_estab_edu	Educational services
ann_payK_finance	no_estab_finance	Finance and insurance
ann_payK_health	no_estab_health	Health care and social assistance
ann_payK_info	no_estab_info	Information
ann_payK_manage	no_estab_manage	Management of companies and enterprises
ann_payK_manufac	no_estab_manufac	Manufacturing
ann_payK_mining	no_estab_mining	Mining, quarrying, and oil and gas extraction
ann_payK_nc	no_estab_nc	Industries not classified
ann_payK_other	no_estab_other	Other services (except public administration)
ann_payK_prof_tech	no_estab_prof_tech	Professional, scientific, and technical services
ann_payK_re	no_estab_re	Real estate and rental and leasing
ann_payK_retail_trade	no_estab_retail_trade	Retail trade
ann_payK_services	no_estab_services	Accommodation and food services
ann_payK_total	no_estab_total	Total for all sectors
ann_payK_transport	no_estab_transport	Transportation and warehousing
ann_payK_utilities	no_estab_utilities	Utilities
ann_payK_whole_trade	no_estab_whole_trade	Wholesale trade

Table 10: Data from County Business Patterns, US Census Bureau

year	GDP per capita	GDP percent change	County and State
2016	39.39	100.58	Brunswick Virginia
2016	435.04	93.28	Hutchinson Texas
2016	35799.74	78.38	Loving Texas
2016	119.2	77.66	Steele North Dakota
2016	63.07	52.07	Briscoe Texas
2016	175.21	45.57	Monroe Ohio
2016	76.81	36.84	Campbell South Dakota
2016	84.89	35.86	Griggs North Dakota
2016	64.13	34.35	Emmons North Dakota
2016	61.48	34.02	Renville North Dakota
2017	90.83	61.22	Adams North Dakota
2017	557.96	59.83	Reeves Texas
2017	289.02	58.42	Storey Nevada
2017	61.26	49.87	Perkins South Dakota
2017	69.9	49.64	Wibaux Montana
2017	169.51	45.05	Winkler Texas
2017	115.35	42.25	Clark Kansas
2017	140.71	39.05	Howard Texas
2017	80.53	37.03	McIntosh North Dakota
2017	57.14	35.7	Floyd Texas
2018	47.4	87.82	Jackson West Virginia
2018	58.09	72.57	Harlan Nebraska
2018	93.76	65.37	Banner Nebraska
2018	36.55	61.14	Chouteau Montana
2018	63.06	58.39	Arthur Nebraska
2018	69.16	52.16	Lamoure North Dakota
2018	42.26	50.31	Sherman Nebraska
2018	78.81	47.74	Traverse Minnesota
2018	76.68	47.66	Jones South Dakota
2018	807.04	44.64	Reeves Texas

Table 11: Top 10 Counties by GDP percent change and year

year	GDP per capita	GDP percent change	County and State
2016	58.39	-28.11	Claiborne Mississippi
2016	45.17	-29	Kidder North Dakota
2016	25.48	-29.08	Benson North Dakota
2016	32.99	-29.91	Harmon Oklahoma
2016	28.76	-30.17	Taylor Iowa
2016	77.35	-30.7	Castro Texas
2016	44.67	-33.19	Arthur Nebraska
2016	46.55	-33.42	Haakon South Dakota
2016	95.34	-35.24	Divide North Dakota
2016	67.05	-37.07	Petroleum Montana
2017	16.3	-30.11	Mora New Mexico
2017	48.87	-30.62	Grant Nebraska
2017	39.93	-31.36	Linn Kansas
2017	35.37	-32.9	Clark South Dakota
2017	65.07	-32.92	Slope North Dakota
2017	35.01	-33.83	Sheridan North Dakota
2017	73.95	-34.21	Sioux Nebraska
2017	45.45	-39.3	Lamoure North Dakota
2017	48.94	-40.44	Hayes Nebraska
2017	56.69	-53.66	Banner Nebraska
2018	49.32	-25.86	Daniels Montana
2018	30.17	-30.32	Mercer Missouri
2018	25.17	-33.16	Milam Texas
2018	102.29	-34.08	Sherman Texas
2018	58.91	-34.13	Loup Nebraska
2018	30.13	-34.85	Adams Ohio
2018	57.21	-36.98	Castro Texas
2018	55.46	-39.4	Rock Nebraska
2018	49.59	-42.88	Blaine Nebraska
2018	34.65	-43.95	Grant North Dakota

Table 12: Bottom 10 Counties by GDP percent change and year

List of Figures

List of Tables

References

[1] Housel, Morgan. "How This All Happened." *The Collaborative Fund*, 14 Nov 2018, https://www.collaborativefund.com/blog/how-this-all-happened/

[2] Judis, John B. "It's the Economies, Stupid." *The Washington Post Magazine*, 29 Nov 2018, https://www.washingtonpost.com/news/magazine/wp/2018/11/29/feature/the-key-to-understanding-americas-red-blue-split-isnt-ideology-or-culture-its-economics/?noredirect=on&utm_term=.0fc33ca9ee52

[3] Bureau of Economic Analysis, https://www.bea.gov

[4] US Census Bureau, https://www.census.gov

[5] US Bureau of Labor Statistics, https://www.bls.gov

[6] Kuhn, Max and Johnson, Kjell. (2013) *Applied Predictive Modeling*. New York: Springer.