

Oilonomics: Making sense of the impact of hydrocarbon production on socio-economics

Farrukh Hamza¹, Qian Sun¹, Abdunnaser Erwemi¹, Jin Haeng Lee²,
Ashfak Ahmed Chowdhury³, Quyen Vu Nguyen⁴

¹Georgia Institute of Technology, North Avenue NW, Atlanta, GA, USA

²Borehole Seismic LLC, 10500 Northwest Freeway, Suite 224, Houston, Texas USA

³General Motors, 300 Renaissance Ctr Detroit, MI USA

⁴Bright Health, 219 North 2nd Street, Minneapolis, MN USA

fhamza@gatech.edu, qian.sun@gatech.edu, aerwemi3@gatech.edu, jinhaeng.lee.87@gmail.com,
asif.chowdhury64@gmail.com, quenvunguyen@gmail.com

Abstract— The development of the oil and gas industry has stimulated local economic growth by creating employment opportunities, attracting migrants, and increasing local business investments. While previous research shows controversial results of the oil and gas industry's influence on the local society, very few studies concentrated on visualizing the effect of such influences and predicting the trends to serve business, local government, job seekers, and the public. This study demonstrates correlations between oil/gas production and the socio-economic variables over time with multiple massive data sets extracted from official websites. Our initial attempt is to visualize such correlations and trends in Colorado while possibly generalize to the national and even international levels for further study.

Keywords— Petroleum Industry, Employment Opportunities, Oil/Gas, Socio-economics

I. INTRODUCTION

In a recent survey of 1200 people by Ernst & Young [1], the public perception of the energy industry was “positive” by only 40% of the participants, described by Ernst & Young as “respectable but precarious.” The perceptions matter because it affects the ability of the oil and gas (O&G) industry to recruit and retain talent, raise capital, develop projects, influence tax and regulation, and maintain its long-term viability, as well as influence public policy [40]. Similarly, in other studies, the researchers [2, 3, 4, 5, 6, 7] mentioned a few reasons for public mistrust and stressed the need to highlight the economic benefits and provide stakeholders with visibility. Public engagement is essential to ensure the sustainable development of energy resources.

The O&G industry boom has stimulated local economic growth [10] by creating employment opportunities, attracting migrants, and increasing local business investments [11, 12, 13]. Examples include North Dakota, Texas, Alberta, Louisiana, Colorado, New Mexico, and Oklahoma. A 2011 New Mexico study [8] found a positive effect on employment,

income, and population, while a Pennsylvania study [9] found negligible income impacts on the general population. A Permian Basin study [43] found significant spillover economic effects to surrounding regions and indirect industries. However, it can also cause a strain the local resources and has potential risks related to public health [44]. As discussed above, several researchers from many parts of the world have discussed how expansion in oil production has directly affected the country's/region's socio-economic growth [22, 23, 24, 25]. Nevertheless, no researcher has created any interactive visualizations for the readers to observe this concept using a combination of oil production data, census data, and employment data. Our project attempts to accomplish the shortcoming. Therefore, to help various stakeholders to understand the O&G industry and provide an unbiased view of data for policymakers, investors, job seekers, and the public, our team has attempted to visualize the data of O&G production while overlapping with socio-economic information using Colorado as an example. We extracted the county-level O&G production data from the Colorado Oil and Gas Conservation Commission (COGCC) [14]. Our approach is to merge this dataset with socio-economic data such as poverty, population, and employment record [15]. In this project, we attempt to demonstrate correlations between oil/gas production and the socio-economic variables over time.

II. SURVEY

Previous research has shown that population growth due to the development of O&G can be rapid [26, 28]. Oil and gas companies were also able to forge close working relationships with state regulators who believed that building and maintaining a strong economic base offered multiple benefits for the state [27]. In 2009, Colorado ranked 6th in US natural gas production. According to IHS Global Insight (2009), the natural gas industry employs over 137,000 people in Colorado directly or indirectly, accounting for approximately 6% of its workforce [27]. However, researchers also found that public health issues are a concern. For example, populations living

near O&G development may experience potential rural, economic, participatory, or distributive injustices and may bear greater vulnerability to risks associated with O&G development than the general population [26].

The O&G sector has reviewed the technologies of Industry 4.0 [35], which offers energy saving and environmental protection as one of the potential benefits. Visualization has been a vital component of the workflows of the O&G industry, such as visualizing subsurface 3D seismic and reservoir data [36, 37]. The study in [32] provided an introduction of visualization techniques for non-experts in O&G. However, its focus is on the surface, and subsurface characterization for offshore data (Gulf of Mexico) is beyond the scope of our project. Recent research [33] has focused on big data and modern computing technologies, providing high-level ideas and insights for data analytics workflow in O&G companies, but they do not provide actual examples. A detailed description of real-time data processing and user interaction tools is available [34]—however, none of these researchers [32, 33, 34] use socio-economic data.

More recently, there have been several O&G production data visualization websites provided by companies such as Enervus [16], RS Energy [17], ShaleProfile [18], and IHS Markit [19]. However, their focus is O&G production, and hardly any visualizations involve socio-economic data. Besides, these subscriptions are proprietary and expensive, thus fail to provide access to the general public. Visualizations provided by websites such as [20, 21] are free but with a propagandistic view of oil and gas data. Therefore, our visualization tool provides a much-needed service because we include multiple datasets and aim to show the correlations of the selected variables over time.

Proposed Method

A. Our Technical Approach

The team registered a domain to serve the visualizations (<http://www.theoilonomics.com>). The technology stack we have employed includes Azure offerings. The significant components are virtual machine instances (Linux platform) and CosmosDB, an emulation of MongoDB for Azure, stored on a distributed cluster. We are using GitHub (Georgia Tech student account) to share code and version control and host the Tableau online app's current version.

We obtained detailed data from the COGCC website. To survey the oil production data, we extracted the well's identifications (API numbers) and their locations, based on COGCC's website. Based on this list of wells, we then queried the COGCC reports for each well per month, parsed the returned HTML site for production data. We then queried the cloud database to check if such a record already exists, and populated the record. The data constitutes more than 120,000 wells and 20 million data points. For our local machine data exploration, we have used SQLite and QGIS to understand and visualize a subset of the data. Due to unknown and

changing schema, we decided to use MongoDB, which is less sensitive to a pre-defined schema.

Overall, the addition of a cloud database results in expensive CRUD operations. Even though it allows freedom to create a record, indexing and querying database are inefficient. The Cloud database introduced priced request unit cap and timeout limits, unnecessary for local databases. Data aggregation is also an abstraction of query calls which were not set up in our cloud database but leverage some form of caching/hashing. Therefore, the non-optimized database required the queries and aggregations to be optimized externally or risked exceeding timeout limits. We ended up using our SQLite database, and files cleaned using Jupyter notebook scripts to populate our analyses that set up in Tableau.

B. Design of Our Approaches

The design of our approaches has been proved to be useful for stakeholders. Energy is a trending topic in policymaking in modern times, and the public is interested in understanding the O&G industry. Policymakers are eager to learn the available data and how it relates to the local economy. Investors, businesses, and job seekers urge to have an unbiased view of the data. Discovered patterns and presented socio-economic development related to the O&G industry can be used in projecting future development [31]. A water usage report [38] related to unconventional shale production in Colorado highlights that the public needs to be informed on acceptable trade-offs [39], thus diminishing conflicts.

A successful “Oilonomics” project can allow the State of Colorado to make educated policies by mitigating negative impacts and maximizing economic potential [29], which benefits both the public and the O&G industry. If it is successful, this tool can be used at national or international levels. The success case can be an interactive web-based visualization that allows access of the O&G production and socio-economic factors. We can also measure this success by user studies, user visits, and interaction statistics at the website.

There were some risks in this project. Our data has some issues such as incompleteness or missing values [41,42 discuss various ways of working with data having missing values]. In some instances, we had to change our approach because too much or too little data to draw a valid conclusion. Technical implementation proved to be challenging and time-consuming. Combining different datasets collected by different agencies at different scales [30] was difficult.

We synthesized datasets extracted from the United States Department of Agriculture (USDA) Economic Research Service, Bureau of Economic Analysis, and Colorado Information Marketplace. Apart from running regressions to explore the impacts of the oil/gas industry on various socio-economic factors, we also examined how the associations transformed along with the time within one or two decades. We further explained our approaches for each experiment in

the next section. The user interface is based on Tableau dashboards (shown at <http://www.theoilonomics.com>) and came about multiple iterations. The user interface's goal is to let users interact with the data on the home page while making our analyses available to them in simplified interactive dashboards. The dashboard joins multiple data sources to represent the wells' location on a satellite map which works as a filter for all the charts such as Regression, Time Series & Area Charts.

We adopted the methods of principal component analysis (PCA), linear regression, and random forest to examine the associations for the four hypotheses. We used Python for data analyses and provided associated files in the CODE folder. Based on the availability of production data, 38 Colorado counties are involved in running the models. We adopted the BOE (barrel of oil equivalent) metric to calculate the total oil/gas production and oil value ($=\text{BOE} \times \text{historical oil price of the year}$) to synthesize oil and gas production. We also calculate the change in oil production (oil diff) from 2010-2018 for model prediction.

Intuition and List of Innovations: Our solution is better than the existing tools because no such public tool currently exists and comprises the following innovations.

- Examining the correlations between oil/gas production information (production, location, and so on) and socio-economic factors with multiple datasets.
- Interactive visualization of oil/gas and all possible socio-economic factors in the same visualization interface.
- Exploring the trend/correlation over time (i.e., time-series) based on historical data of the oil/gas industry.
- Open source tech stack for visualization of oil/gas and socio-economics (Enervus [16], RS Energy [17], IHS Markit [19], and more are commercial oil/gas vendors with proprietary data streams and tech stack).

III. EXPERIMENTS AND EVALUATION

A. Description of Testbed

The boom of the oil industry has attracted migration, created employment, and enhanced local economics. Employment and unemployment rates are much-discussed attributes of the socio-economy concerning hydrocarbon production. With advancing technology allowing remote access of workflows and reduced project length, the impact is on direct employment, indirectly attributable jobs, and non-related jobs. As a global powerhouse, oil and gas also impact the macro economies. Implying causation would be difficult since there are many possible external factors such as economics and fiscal policies. Therefore, we propose to test the following hypotheses, using appropriate testing metrics and datasets.

1. Migration is associated with oil/gas production.

2. An increase in oil production will also increase the employment rate over time.

3. Oil production is correlated with indirectly attributable jobs as well as non-oil/gas-related employment.

4. Increase in oil production increases local trade volume.

B. Details of Experiments

1) Hypothesis: migration is associated with oil/gas production. To examine the association of migration and oil/gas production, we selected estimated total population, oil value, change of oil production, and natural population increase as top features that associate to the population migration. Among them, the county-level estimated total population and natural population increase (birth-death) are control variables. The estimated total population in Colorado grew every year, but the population migration varied with different counties. In general, we found that the selected features do not show associations with domestic migration and net migration (the difference between international migration and domestic migration). However, the independent variables demonstrate associations with international migration. In the attribute explained variance ratio, PCA's results indicated that the first component contains 50.71% of the variance. The second principal component contains 34.20% of the variance, and the third principal component contains 12.69% of the variance. Together, the three components contain 97.60% of the information. Despite the collinearity of population and natural increase, the linear regression model illustrates that both variables of oil value and the difference in oil production show statistically significant correlations with international migration. The model has an adjusted R-square of 0.845. The results of random forest also indicate the existing associations between oil value, the annual change of oil production, and international migration. The mean absolute error of random forest is 107.96, indicating 83.36% $((648.81 - 107.96)/648.81)$ improvement over the baseline error.

As mentioned above, migration varied with counties in Colorado. Therefore, we used the same model to test some individual counties with the most synthesized oil production (BOE), including Weld, Garfield, La Plata, Rio Blanco, and Las Animas. With the same selection of dependent and independent variables, we found that for the linear regression model, the annual oil change demonstrates a statistically positive correlation with international migration. In contrast, the oil value did not have a significant correlation. The random forest results illustrate that BOE shares a similar weight associating with international migration (15%) compared to the total estimated population (17%), with the root mean square error of 24.8. Therefore, the results support our hypothesis that migration and oil/gas production is associated.

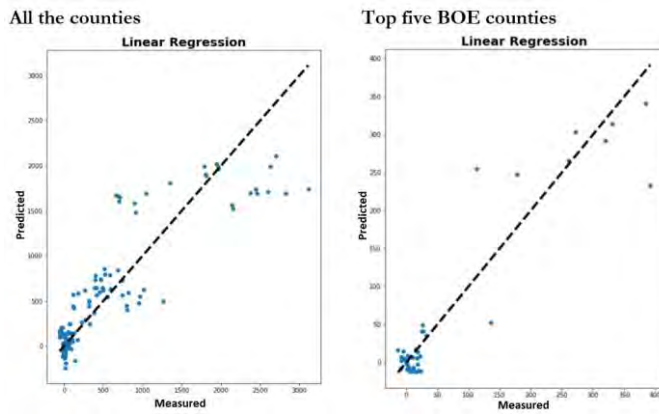


Figure 1. Linear Regression result using all the counties (left), versus only the top 5 producing counties shows that our hypothesis of migration and oil/gas production are associated. The x-axis label is “Measured”, while the y-axis label is “Predicted”.

2) Hypothesis: an increase in oil production will also increase the direct employment rate over time. To investigate the hypothesis about the employment rate of jobs that are directly related to oil and gas, we selected oil sales, gas sales, oil production difference, wage and salary, and BOE as features, with oil industry employment being the response variable. The correlation matrix [figure 2.1] suggests that all features are correlated with one another except BOE. PCA demonstrates that one principal component is predominant in terms of importance, consistent with LASSO result in that ‘Wage and salary’ dominate the importance. Ordinary least squares (OLS) was performed and resulted in an R-square value of 0.864, providing a good fit but no apparent connection to accept the hypothesis. Finally, the random forest was performed and resulted in 12% mean absolute error. We also executed variable importance, which generated similar results to PCA and LASSO, with ‘wage and salary’ to be the dominant feature. To further analyse how BOE affects employment, we performed Pearson’s linear correlation and Spearman’s nonlinear correlation. Accounting for historically known oil-producing counties (Weld, Arapahoe, Lincoln, Cheyenne, Rio Blanco, Garfield), and gas-producing counties (Weld, Rio Blanco, Garfield, Dolores, Las Animas, Montezuma, La Plata), all oil-producing counties and the majority of gas-producing counties are positively correlated (figure 2.1). Their degree of correlation was related to oil production or BOE. We found no meaningful correlation between BOE of all counties and direct employment for each year during 2010-2018. We conclude a robust correlation between direct employment and features mentioned above; however, mostly due to total wage and salary, not because of BOE increase/decrease. According to our result, an increase in wage and salary might have caused an increase in employment, and thus further investigation is needed to observe the relationship between total wage and salary with BOE.

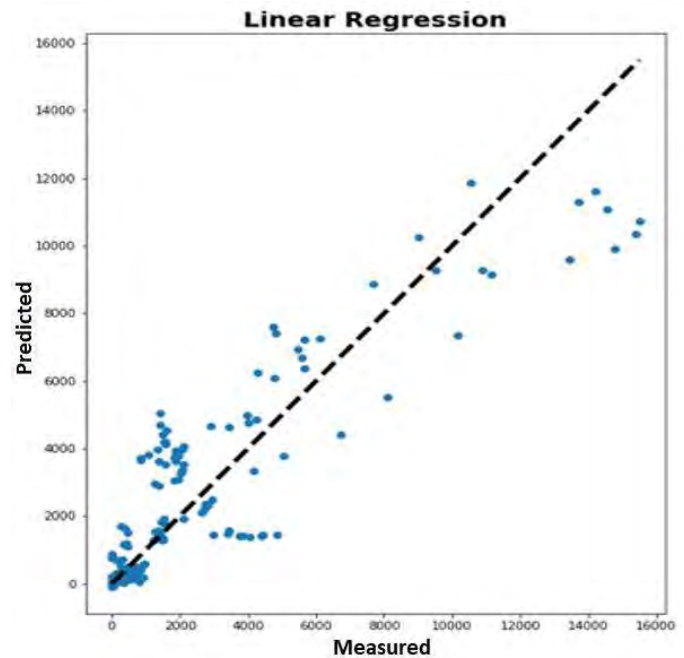


Figure 2. Linear Regression Result with direct employment as response variable. The x-axis label is “Measured”, while the y-axis label is “Predicted”.

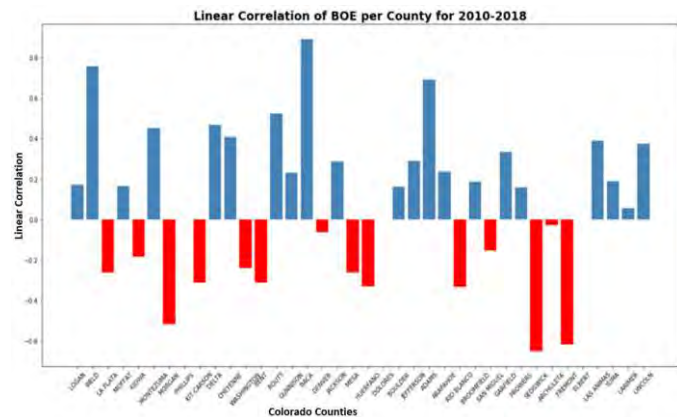


Figure 3. Pearson’s Linear Correlation of BOE with direct employment for all Counties throughout year 2010-2018. The x-axis labels are Colorado county names, while the y-axis label is “Linear Correlation”.

3) Hypothesis: oil production is correlated with indirectly attributable jobs as well as non-oil/gas-related employment. Features used here are totwage (total wage), estab (the number of employer establishments in the industry), avgwk wage (average weekly wage per worker), and BOE, with the response variable being avgemp (number of workers employed in the industry). As can be observed [Figure 3.1], features are significantly correlated. We performed LASSO and PCA, the results indicate one feature explaining most of the variance. In contrast, BOE shows very weak importance. We performed the OLS with an R-square of 0.994. The overfit model might be the result of collinearity. Additionally, we used the random forest as a secondary analysis, producing only 3% of the mean absolute error. Nevertheless, when observing variable importance, the importance of BOE was negligibly low. Given prior information on primary oil-

producing counties (Weld, Arapahoe, Lincoln, Cheyenne, Rio Blanco, Garfield), and gas-producing counties (Weld, Rio Blanco, Garfield, Dolores, Las Animas, Montezuma, La Plata), results observed from the linear correlation bar graphs show both positive and negative correlations, indicating no exact way to accept the hypothesis. We speculate that the county needs to produce an amount of oil and gas for a significant increase in employment, exemplified in Weld County. Linear correlation throughout years 2010-2018 for all counties seem to follow oil price trend when falling and rising (note the oil crisis took place in 2014-2016), yet when there is no significant change, we saw a negative correlation. In conclusion, we cannot accept the hypothesis without further investigation.

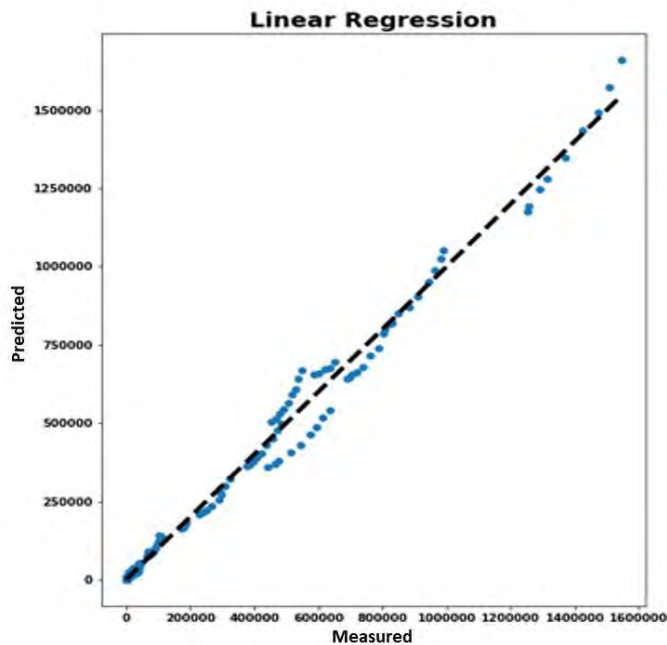


Figure 4. Linear Regression result with indirect employment as response variable. The x-axis label is “Measured”, while the y-axis label is “Predicted”.

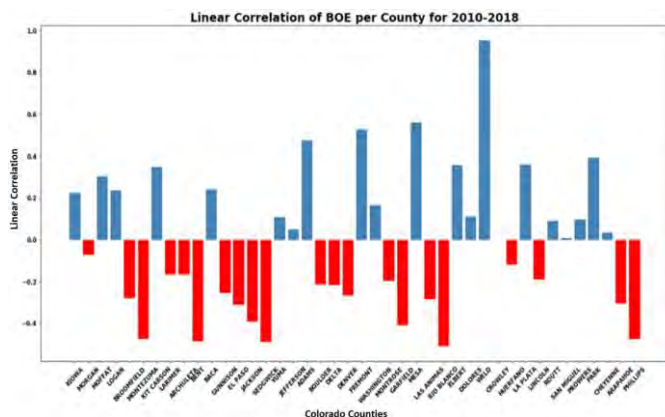


Figure 5. Pearson's' Linear Correlation of BOE with indirect employment for all Counties throughout year 2010-2018. The x-axis labels are Colorado county names, while the y-axis label is “Linear Correlation”.

4) Hypothesis: an increase in oil production increases local trade volume. Similar to the association with the employment rate, oil production might also contribute to other

local economies. With the data at the state level, we tried to explore the associations between BOE production and all sorts of expenditures.

First, we examined the Pearson correlation of BOE, nondurable goods, personal consumption expenditures, household consumption expenditures (for services), durable goods, and final consumption expenditures of non-profit institutions serving households. For BOE, we found a strong correlation with all the selected features, among which nondurable goods, personal consumption expenditures, household consumption expenditures (for services), and final consumption expenditures of non-profit institutions serving households reached more than 90% of the correlation. In the model, we included BOE, nondurable goods, household consumption expenditures (for services), and wage and salary employment after trials and experiments.

The features we selected demonstrate some associations with furnishings and durable household equipment. PCA results suggest that the first principal component has already taken up 96.33% of the variance. The best score using built-in LASSO cross-validation of 0.92. The model of linear regression appeared overfit with adjusted R-square to be 0.996. The model also fails to show any statistically significant correlation between oil production (BOE) and the dependent variable (furnishings and durable household equipment). Based on the results of random forest, we found that the furnishings and durable household equipment are top drivers (24%), surpassing that of the nondurable goods (17%). We did not find concrete evidence to support this hypothesis.

IV. CONCLUSIONS AND DISCUSSION

We successfully delivered our results at theoilonomics.com website to disseminate information in an interactive and accessible fashion. Our four hypotheses yielded the following results:

- International migration is associated with oil production (BOE)/oil value, yet the direction of association varied across counties. No statistically significant correlation between domestic migration and oil production/oil value was found.
- Through various analyses, we conclude that direct employment is correlated with BOE, and the degree of correlation is related to the degree of BOE for oil producing counties. Thus, we accept the hypothesis.
- We found that the variable importance of BOE was low, along with other features that showed strong collinearity. The strong nonlinear and linear correlation seems to hold for counties that produced oil/gas over a certain amount. Hence without further investigation, we cannot accept the hypothesis.
- We are expecting the local expenditures would increase along with the rise of oil production. Due to the limited data, we did not find solid evidence to support our hypothesis. However, the results of the Pearson Correlation show promising potential for further investigation of this hypothesis.

ACKNOWLEDGMENT

We, the authors, want to show our gratitude to Duen Horng (Polo) Chau, Associate Professor at Georgia Institute of Technology, and Maksym Pryporov, Mathematical Scientist at Occidental Petroleum Corporation, for sharing their wisdom to improve this research study. We thank three anonymous reviewers for their insights which helped to improve the manuscript.

REFERENCES

- [1] Rassenfoss, S. (2019, March 1). The Challenge of Public Perception. Society of Petroleum Engineers. doi:10.2118/0319-0027-JPT
- [2] Edman, S. (2013, June 1). Public Perception of the Oil and Gas Industry. Society of Petroleum Engineers. doi:10.2118/0213-015-TWA
- [3] Theodori, G. L. (2008, January 1). Public Perception of the Natural Gas Industry: Insights from Two Barnett Shale Counties. Society of Petroleum Engineers. doi:10.2118/115917-MS
- [4] Theodori, G. L., & Jackson-Smith, D. (2010, January 1). Public Perception of the Oil and Gas Industry: The Good, the Bad, and the Ugly. Society of Petroleum Engineers. doi:10.2118/134253-MS
- [5] Theodori, G. L. (2008, January 1). Public Perception of the Natural Gas Industry: Insights from Two Barnett Shale Counties. Society of Petroleum Engineers. doi:10.2118/115917-MS
- [6] Newport, F. 2007. Annual Update: Americans Rate Business and Industry Sectors: Oil and gas industry and federal government have worst images. . Downloaded Feb 26, 2020.
- [7] Higgins, M., Burnett, D., Kreuter, U., Theodori, G., & Haut, R. (2018, September 24). Public Engagement: New Approaches to Enhance Field Planning and Environmental Compliance for O & G Development. Society of Petroleum Engineers. doi:10.2118/191739-MS
- [8] Peach, J., & Starbuck, C. M. (2011). Oil and Gas Production and Economic Growth in New Mexico. *Journal of Economic Issues*, 45(2), 511–526. doi:10.2753/jei0021-3624450228
- [9] Paredes, D., Komarek, T., & Loveridge, S. (2015). Income and employment effects of shale gas extraction windfalls: Evidence from the Marcellus region. *Energy Economics*, 47, 112–120. doi:10.1016/j.eneco.2014.09.025
- [10] Munasib, A., & Rickman, D. S. (2015). Regional economic impacts of the shale gas and tight oil boom: A synthetic control analysis. *Regional Science and Urban Economics*, 50, 1–17. doi:10.1016/j.regsciurbeco.2014.10.006
- [11] Weber, J. G. (2012). The effects of a natural gas boom on employment and income in Colorado, Texas, and Wyoming. *Energy Economics* 34 (5), 1580–1588.
- [12] Corden, W.M., Neary, J.P., 1982. Booming Sector and De-Industrialisation in a Small Open Economy. *Econ. J.* 92, 825–848.
- [13] Feyrer, J., Mansur, E. T., and Sacerdote, B. (April 2017). “Geographic Dispersion of Economic Shocks: Evidence from the Fracking Revolution.” *American Economic Review* 107, no. 4: 1313–34. .
- [14] Colorado Oil and Gas Conservation Commission (COGCC) GISOnline. (n.d.). Retrieved from https://cogccmap.state.co.us/cogcc_gis_online/
- [15] County-level Data Sets. (2020, February 5). Retrieved February 20, 2020, from <https://www.ers.usda.gov/data-products/county-level-data-sets>
- [16] Enverus home page. (2020, February 25). Retrieved February 26, 2020, from <https://www.enverus.com/>
- [17] Technology that Inspires Innovation. (n.d.). Retrieved February 25, 2020, from <https://www.rseg.com/>
- [18] Shale Oil and Gas Production & Completion Data. (n.d.). Retrieved February 20, 2020, from <https://shaleprofile.com/>
- [19] Leading Source of Critical Information. (n.d.). Retrieved February 25, 2020, from <https://ihsmarkit.com/index.html>
- [20] Oil and Gas Threat Map. (2016, March 31). Retrieved February 22, 2020, from <https://oilandgasthreatmap.com/>
- [21] Oil and Gas by Location. (2020, February 24). Retrieved February 26, 2020, from <https://www.fractracker.org/map/>
- [22] Nanok, Josephat Koli; Onyango, Christopher Ouma (2017) : A socioeconomic and environmental analysis of the effects of oil exploration on the local community in Lokichar, Turkana County, Kenya, *International Journal of Management, Economics and Social Sciences (IJMESS)*, ISSN 2304-1366, IJMESS International Publishers, Jersey City, NJ, Vol. 6, Iss. 3, pp. 144-156, <https://www.econstor.eu/bitstream/10419/171451/1/1006098852.pdf>
- [23] PWC; July 2017: Impacts of the Oil and Natural Gas Industry on the US Economy <https://www.api.org/~media/Files/Policy/Jobs/Oil-and-Gas-2015-Economic-Impacts-Final-Cover-07-17-2017.pdf>
- [24] L V Eder; 2018: Exploration and production of oil and gas as a factor of economic development of the Arctic resource regions, resource curse
- [25] BBC Research & Consulting; July 2013: Social Impacts of Oil and Gas Development on Eastern Montana Communities
- [26] Lisa M. McKenzie, William B. Allshouse, Troy Burke Benjamin D. Blair, and John L. Adgate. 2016. Population Size, Growth, and Environmental Justice Near Oil and Gas Wells in Colorado. *American Chemical Society*
- [27] Charles Davis. 2012. The Politics of “Fracking”: Regulating Natural Gas Drilling Practices in Colorado and Texas. *Review of Policy Research*, Volume 29, Number 2 (2012)
- [28] Mark Agerton, Peter R. Hartley, Kenneth B. Medlock III, Ted Temzelides. 2017 Employment impacts of upstream oil and gas investment in the United States. *Energy Economics* 62 (2017)(171–180)
- [29] Obidzinski, K., R. Andriani, H. Komarudin, and A. Andrianto. 2012. Environmental and social impacts of oil palm plantations and their implications for biofuel production in Indonesia. *Ecology and Society* 17(1): 25.
- [30] Cong R., Saito M., Hirata R., Ito A., Maksyutov S. 2018. Visualization on fossil-fuel carbon dioxide (CO₂) emissions from buildings in Tokyo metropolis. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4 (4) , pp. 21-25.
- [31] Winch, Peter,M.D., M.P.H., & Stepnitz, R., M.H.S. (2011). Peak oil and health in low- and middle-income countries: Impacts and potential responses. *American Journal of Public Health*,
- [32] Slatt, R. M., Thomasson, M. R., Romig, Jr., P. R., et al. 1996. Visualization Technology for the Oil and Gas Industry: Today and Tomorrow. *AAPG Bulletin*, V. 80, No. 4 (April 1996), P. 453–459.
- [33] Mohammadpoor, M., Torabi, F. 2018. Big Data analytics in oil and gas industry: An emerging trend. *Petroleum*. <https://doi.org/10.1016/j.petlm.2018.11.001>
- [34] Alkiyumi, S., Kumar, P., Dattana V. 2018. Digital Oilfield: Real-Time Production Data Visualizations (Wells-Testing Operation). *RESEARCH REVIEW International Journal of Multidisciplinary*. Vol-3 | Issue-08 | August 2018 | Published Online: 07 August 2018.
- [35] H. Lu, L. Guo, M. Azimi, and K. Huang, “Oil and gas 4.0 era: A systematic review and outlook,” *Computers in Industry*, vol. 111, pp. 68–90, 2019.
- [36] Wolfe, R. H. and Liu, C. N., “Interactive Visualization of 3D Seismic Data: A Volumetric Method”, *IEEE Computer Graphics and Applications*, 8, 7 (July 1988), 24–30.
- [37] Sultanum, N., Somanath, S., Sharlin, E., Sousa, M. C. “Point it, Split it, Peel it, View it”: Techniques for Interactive Reservoir Visualization on Tabletops. In *Proc. ITS* (2011), ACM, New York, pp. 192–201. doi> 10.1145/2076354.2076390
- [38] Panagiotis D.Oikonomou., Julie A.KallenbergeraReagan M.Waskom., et al. 2016. Water acquisition and use during unconventional oil and gas development and the existing data challenges: Weld and Garfield counties, CO. *Journal of Environmental Management*, Volume 181, 1 October 2016, Pages 36-47
- [39] Bandyopadhyay, G., Bagheri, G., Mann, M. 2006. Reduction of fossil fuel emissions in the USA: A holistic approach towards policy formulation. *Energy Policy*, Volume 35, Issue 2, February 2007, Pages 950-965
- [40] Burstein P. 2003. The impact of public opinion on public policy: a review and an agenda. *Political Research Quarterly*, Vol. 56, No. 1 (March 2003): pp. 29-40
- [41] Templ M, Alfons A, Filzmoser P (2012). “Exploring Incomplete Data Using Visualization Techniques.”*Advances in Data Analysis and Classification*,6, 29–47.doi:10.1007/s11634-011-0102-y

- [45] Kowarik, A., Templ, M. Imputation with the R package VIM. *Journal of Statistical Software*, 74(7):1–16, 2016. doi: 10.18637/jss.v074.i0
- [46] Wang, H. (2020). The economic impact of oil and gas development in the Permian Basin: Local and spillover effects. *Resources Policy*, 66, 101599. doi:10.1016/j.resourpol.2020.101599
- [47] Jackson, R. B., Vengosh, A., Carey, J. W., Davies, R. J., Darrah, T. H., O'Sullivan, F., & Pétron, G. (2014). The Environmental Costs and Benefits of Fracking. *Annual Review of Environment and Resources*, 39(1), 327–362. doi:10.1146/annurev-environ-031113-144051
- [48] S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.
- [49] J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.
- [50] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," *IEEE Electron Device Lett.*, vol. 20, pp. 569–571, Nov. 1999.
- [51] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in *Proc. ECOC'00*, 2000, paper 11.3.4, p. 109.
- [52] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.
- [53] (2002) The IEEE website. [Online]. Available: <http://www.ieee.org/>
- [54] *FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.
- [55] "PDCA12-70 data sheet," Opto Speed SA, Mezzovico, Switzerland.
- [56] A. Kamik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.
- [57] J. Padhye, V. Firoiu, and D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.
- [58] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.
- [59] S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks," *IEEE Trans. Neural Networks*, vol. 4, pp. 570–578, July 1993.
- [60] R. J. Vidmar, On the use of atmospheric plasmas as electromagnetic reflectors. *IEEE Trans. Plasma Sci.* [Online]. 21 (3). pp. 876–880. Available: <http://www.halcyon.com/pub/journals/21ps03-vidmar>, Aug. 1992.



Farrukh Hamza is a graduate student at Georgia Institute of Technology, and works as an Engineering Team Lead at Halliburton, a leading oil and gas services company. His interests include petroleum engineering, hydraulic fracturing, data analytics, and machine learning. He is also involved in technology scouting and contributing to the clean-tech ecosystem. Hamza has authored 15+ industry publications and holds 8 patent applications. He holds a bachelor's degree in mechanical engineering from NED University of Engineering and Technology, and a master's degree in petroleum engineering from the University of Texas at Austin.



Qian Sun graduated from Georgia Institute of Technology with a Master of Analytics. She has broad research interests in social sciences. During her study in the United States, she had participated in multiple research projects involving smallholder farmers in Uganda, the aging population in China, immigrants from Mexico, and intermate partners of homicide cases in the Great Denver Area. Recently she finished a CNX project that applied machine learning models to predict treating pressure using hydraulic fracturing and drilling data. She likes practicing yoga and experiencing different cultures.



Abdunnaser(Naser) Erwemi: A Business Intelligence Developer. Naser has served in several technical and leadership positions in the oil and gas industry. Naser holds a degree in Geophysics and is currently

pursuing a Master in Analytics at Georgia Institute of Technology.



Jin Haeng (John) Lee, from Houston, Texas, was born in Incheon, South Korea. Jin Haeng obtained master's degree in Science with Geophysics major from University of Houston (2016). Jin Haeng is also pursuing to complete his second master's degree (2021) in Analytics with computer science track from Georgia Institute of Technology (Atlanta, Georgia, United States). He is a Geophysicist and a Data Analyst at Borehole Seismic LLC's processing department, located in Houston, Texas. Previously, Jin Haeng served as research assistant with Environmental Engineering research group and Geophysics department from University of Waterloo.



Ashfak Chowdhury (Asif) grew up in Bangladesh and moved to Texas in 2010. He has graduated from University of Texas at Austin with bachelor's degree in Applied Mathematics (2016). He is currently pursuing his master's degree in Business Analytics from Georgia Institute of Technology and working in the automotive industry as a Data Analyst.



Quyen Nguyen is currently pursuing Master of Science in Computer Science at Georgia Institute of Technology. He graduated in 2011 with Bachelor of Science in Chemical Engineering at the University of Texas at Austin, with 7 years of experience in Oil and Gas, but currently is a Software Engineer at Bright Health in Austin, Texas. Quyen was born in Vietnam, is passionate about technology, energy, and data. He also likes tacos.