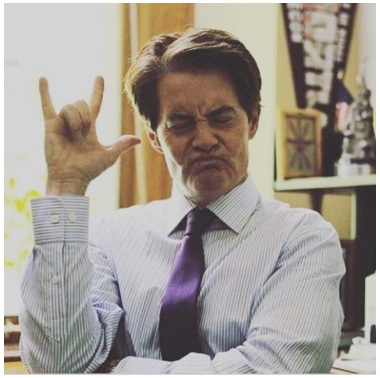
What made the Pacific Northwest move during the July 4th weekend?



Sam Stephens, Eleda Johnson,
Mahesh Arumugam

Berkeley
SCHOOL OF
INFORMATION

Prepared for Mr. Mayor of Portlandia and Oregon Department of Tourism

Introduction

The COVID-19 pandemic has devastated the tourism industry

Good news, is 2021 saw an uptick in domestic travel[2]

In 2020 there was a 35% reduction in domestic travel, resulting in a loss of over 300 Billion dollars[1].

Potential Driving factors:

- ✓ ***Vaccinations***
- ✓ ***Restriction*** *easement*
- ✓ ***Politics***
- ✓ ***New COVID-19 cases***
- ✓ ***Employment***
- ✓ ***Covid fatigue***
- ✓ *Population density*

Research Question

What factors contributed to the decision of embarking on a long road trip (between 100 and 500 miles) during the July 4th, 2021 weekend travel window by the people in Pacific Northwest (WA, OR, ID)?

Data Description

Constraints:

- *Focus on Pacific Northwest (OR, WA, ID)*
- *July 4th weekend, as US holiday shared across region*
- *Statistics take at county level*

Estimator: ***Mobility** - Trips taken between 100 - 500 miles*

Potential Driving factors:

- ✓ ***Vaccinations** - vaccination rate per person*
- ✓ ***Restriction** easement - Indicator variable for active restrictions on businesses*
- ✓ ***Politics** - Trump votes in 2020 presidential race, proxy for conservative views on pandemic*
- ✓ ***New cases** - New cases per 10K people in a 10 day window prior to holiday weekend*
- ✓ ***Employment** - employment rate for June 2021*
- ✓ ***Covid fatigue** - Not operationalized - no survey - omitted variable*
- ✓ *Population density - people from mobility database per square mile of land - Explanatory variable*

Models and Regression

1. Density + one variable directly related to COVID
2. Density + two variables directly related to COVID
3. Density + 'Kitchen Sink':
   - Insight into stability of models 1 and 2
   - Find more significant drivers

	Dependent variable:		
	ln_trips_100_500_per_pop.21		
	(1)	(2)	(3)
ln_pop_density	-0.099**	-0.113***	-0.099**
	(0.031)	(0.030)	(0.038)
ln_vaccinated_first_per_pop	-1.374*	-1.293*	-1.512
	(0.641)	(0.619)	(0.820)
ln_pcases_per_100kpop.21		0.073*	0.071
		(0.035)	(0.056)
ln_employment_rate			10.985*
			(5.296)
rstr_businesses_active			0.331*
			(0.163)
ln_trump_votes_per_pop			0.610
			(0.928)
Constant	-0.442*	-0.699**	-0.522
	(0.201)	(0.221)	(0.551)
Standard Error Type	Robust	Classical	Robust
Observations	119	119	119
Adjusted R^2	0.198	0.220	0.248
Note:	*p<0.05; **p<0.01; ***p<0.001		

Berkeley SCHOOL OF INFORMATION

Model Assumptions

- IID is suspect - regional clustering
- None of the model covariates are perfectly collinear
- Linear conditional expectation are met
- Homoscedasticity assumption met only for Model 2
- All 3 models show normally distributed errors

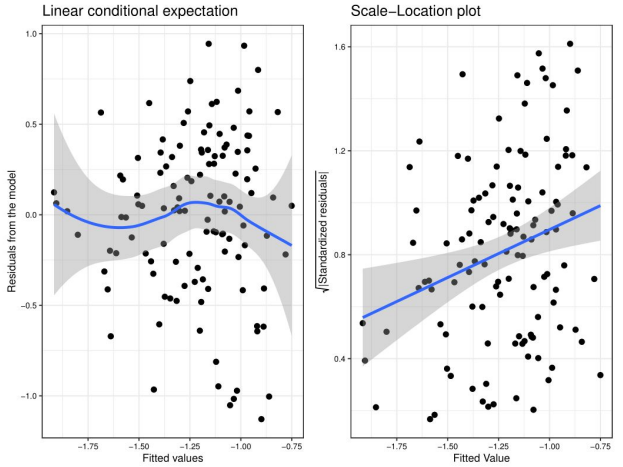


Figure 3: Linear conditional expectation and Homoskedasticity

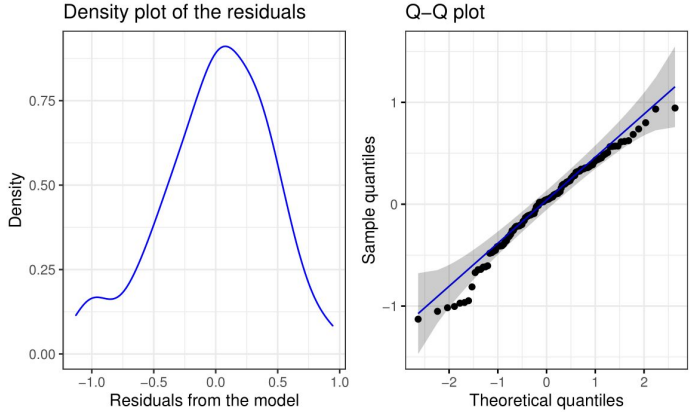


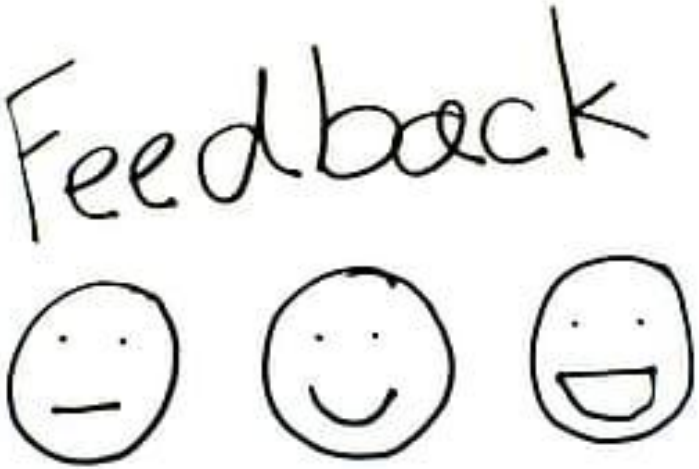
Figure 4: Normally distributed errors

Berkeley SCHOOL OF INFORMATION

Conclusions

	Dependent variable:		
	ln_trips_100_500_per_pop.21		
	(1)	(2)	(3)
ln_pop_density	-0.099**	-0.113***	-0.099**
	(0.031)	(0.030)	(0.038)
ln_vaccinated_first_per_pop	-1.374*	-1.293*	-1.512
	(0.641)	(0.619)	(0.820)
ln_pcases_per_100kpop.21		0.073*	0.071
		(0.035)	(0.056)
ln_employment_rate			10.985*
			(5.296)
rstr_businesses_active			0.331*
			(0.163)
ln_trump_votes_per_pop			0.610
			(0.928)
Constant	-0.442*	-0.699**	-0.522
	(0.201)	(0.221)	(0.551)
Standard Error Type	Robust	Classical	Robust
Observations	119	119	119
Adjusted R^2	0.198	0.220	0.248
Note:	*p<0.05; **p<0.01; ***p<0.001		

Berkeley
SCHOOL OF
INFORMATION


Feedback

EDA Data Pipeline

Initial Cleaning	"The Big Join"	Combined Dataframe
• Standardized variable naming • Calculate trips between 100 and 500 miles • Calculate votes data • Calculate binary restrictions flag • Calculate population • Save interim CSVs	• Join interim CSVs • Bulk calculation of normalizations and transforms	• Search for and resolve by-row anomalies • San Juan County

EDA

- Histograms guided normalization and transformation systems
- Pearson's correlation coefficients guided initial variable selection in model development