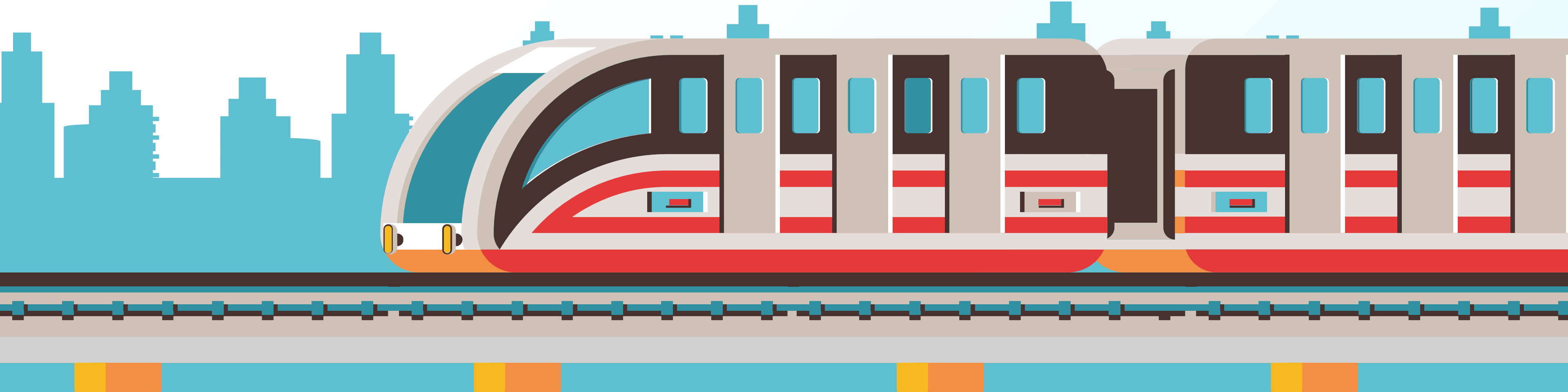
Planes, *Trains*, and Automobiles

Modeling Rail Transportation's Effect on Emissions, Safety, and Travel Time

Saketh Marrapu and Isaac Martinez



Motivation and Introducing the Data

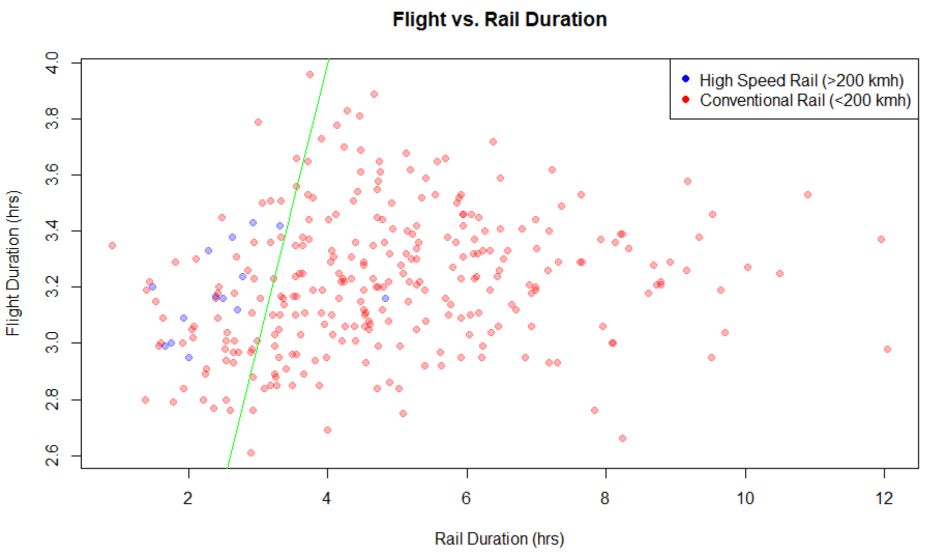
- US is one of the largest countries in the world yet trails other developed countries in its implementation of rail transportation
- There is travel demand for what is largely considered the “goldilocks” zone of rail, about 100-500 miles (“Breakeven Distances”).

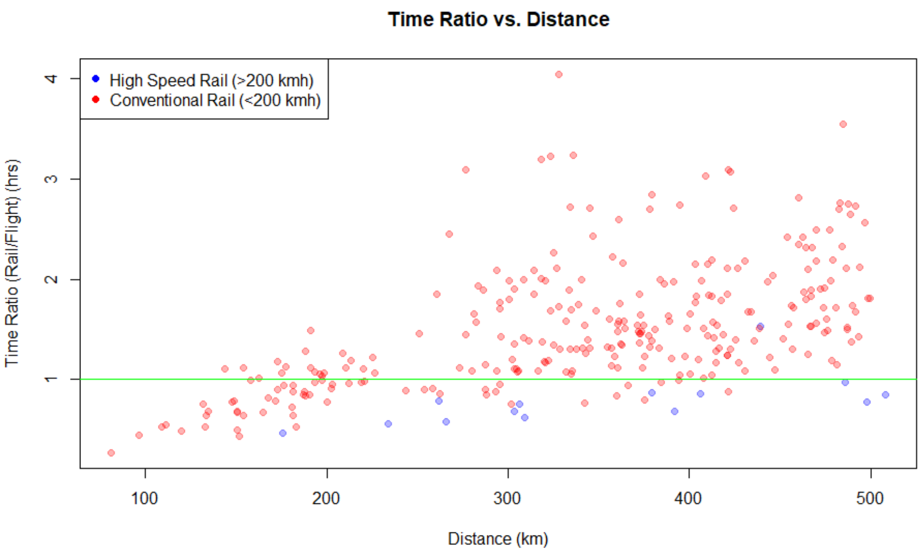
Motivation and Introducing the Data

- Looked at data from the EU, where many countries have established rail networks and reliable data
- With these EU data, our aims were to see rail's effect (relative to air and road transportation) on
  - Travel time
  - Emissions
  - Safety
- After fitting our models, we extrapolate to the US to see how rail *might* measure up against air and road

01

Travel Time





Row	Min.	Median	Mean	Max.	SD
Rail	0.9000000	4.530000	4.793670	12.050000	2.0487048
Flight	2.6100000	3.210000	3.216061	3.960000	0.2420680
Diff	-1.7100000	1.320000	1.577609	8.090000	1.8066368
Ratio	0.2686567	1.391304	1.488632	4.043624	0.6319515

n = 297 (14 HSR, 183 CR)

Paired T-testing

	Conventional	High-Speed
t	14.279	-3.1829
df	282	13
p	< 2.2e-16	0.0072
Diff	1.689	-0.670
CI	(1.456,1.922)	(-1.139,-0.217)

Modeling

Conventional

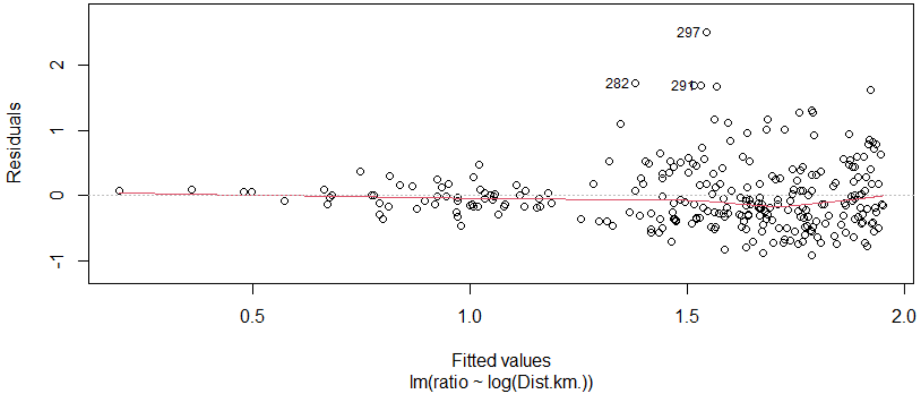
```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -4.0573     0.4714  -8.606 5.45e-16 ***
log(Dist.km.)     0.9668     0.0815  11.863  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5109 on 281 degrees of freedom
Multiple R-squared:  0.3337,    Adjusted R-squared:  0.3313
F-statistic: 140.7 on 1 and 281 DF,  p-value: < 2.2e-16
```

```
        Shapiro-Wilk normality test

data:  resid(crlin)
W = 0.91511, p-value = 1.409e-11
```

Residuals vs Fitted



High-Speed

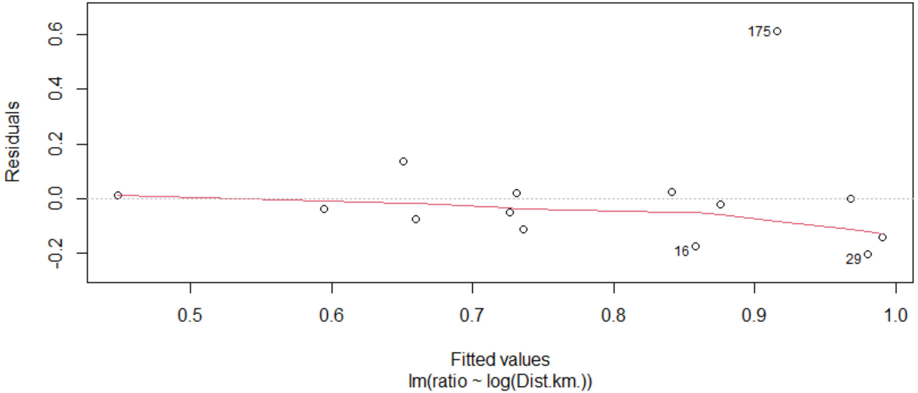
```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -2.1934     1.0533  -2.082   0.0594 .
log(Dist.km.)     0.5110     0.1805   2.831   0.0152 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2058 on 12 degrees of freedom
Multiple R-squared:  0.4004,    Adjusted R-squared:  0.3504
F-statistic: 8.013 on 1 and 12 DF,  p-value: 0.01516
```

```
        Shapiro-Wilk normality test

data:  resid(hsrlin)
W = 0.73284, p-value = 0.0008262
```

Residuals vs Fitted



Extrapolation

State	Route	Dist (km)	Flight Time (hr)	Conventional	High-Speed
TX	DAL-SAT	397	1.167	(1.66,1.80)	(0.73,0.99)
CA	LAX-SFO	543.1	1.5	(1.93,2.13)	(0.80,1.25)
CO	DEN-COS	117.3	0.833	(0.38,0.72)	(-0.19,0.68)
NY-MA	JFK-BOS	300.5	1.333	(1.40,1.52)	(0.59,0.85)

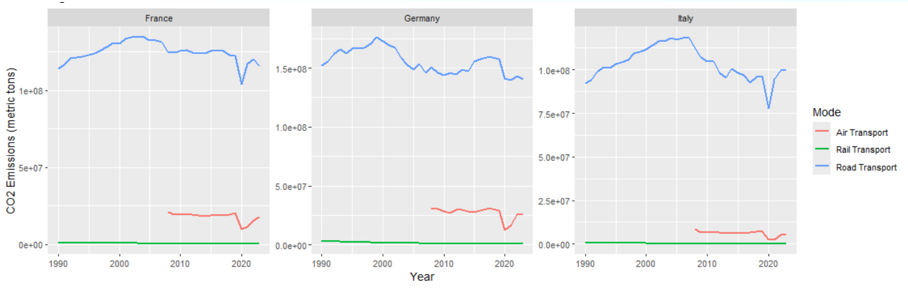
02 Emissions

Emissions: Initial Considerations and EDA

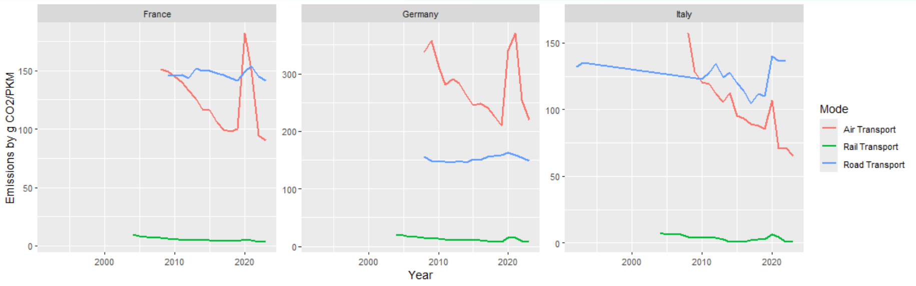
- The data are emissions of CO_2 in metric tons

- Need some way to normalize: passenger-kilometers (PKM) allows to us normalize by usage, giving us g CO_2/PKM

“Raw” Emissions



Normalized Emissions



Emissions: Initial Considerations and EDA

- Turn to t-testing: we test the differences in normalized emissions *per year*

- Years with p-values < 0.05 indicate years when the difference was significantly greater than 0

t-tests of the difference between normalized road and rail emissions, selected years

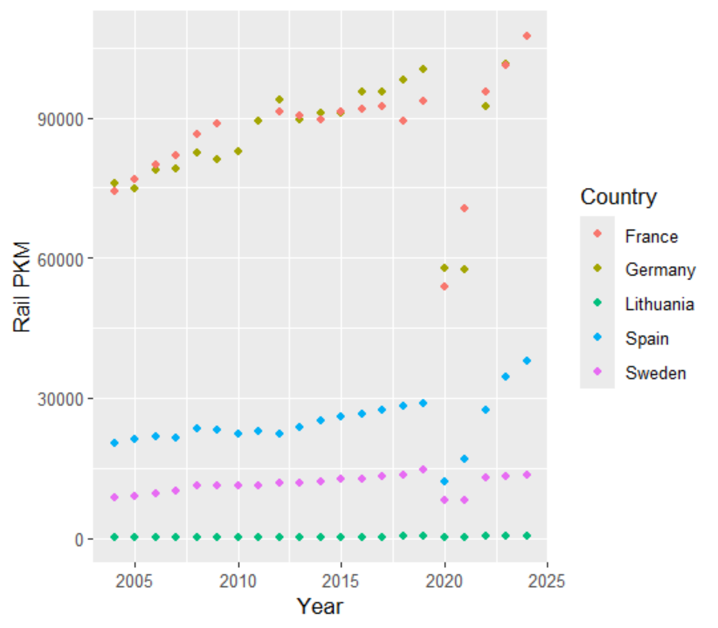
Year	Mean Difference	t-statistic	p-value	n
2008	75.2	0.886	0.198	11
2009	162	1.56	0.0737	12
2010	161	1.58	0.0712	12
2013	99.6	1.9	0.0421	12
2017	89.5	1.53	0.0823	9
2020	140	14.3	0.00000365	7
2023	510	2.15	0.0341	8

t-tests of the difference between normalized air and rail emissions, selected years

Year	Mean Difference	t-statistic	p-value	n
2008	259	1.3	0.103	25
2009	61.9	1.22	0.118	24
2010	291	1.42	0.0842	24
2013	243	1.53	0.0703	24
2017	300	1.56	0.0671	23
2020	930	1.3	0.105	22
2023	336	1.75	0.0471	22

Emissions: Initial Considerations and EDA

- Why are differences not significant?
  - Countries like Lithuania, Estonia, and Croatia consistently give negative difference for one or both differences
  - These countries have lower rail utilization
  - Economic development potentially a factor



Emissions: Modeling

- We start with fitting linear models of both differences year-by-year
- Predictors: GDP per capita, urbanization rate, share of renewable energy
- Hardly any significant, even after simplifying to one predictor, GDP per capita

Single factor (log(GDP per capita)) linear model for Road - Rail for the year 2013

Coefficient estimate	Std. error	p-value
(Intercept) = -25.517224	117.024252	0.832
(GDP) = 0.004020	0.003377	0.261

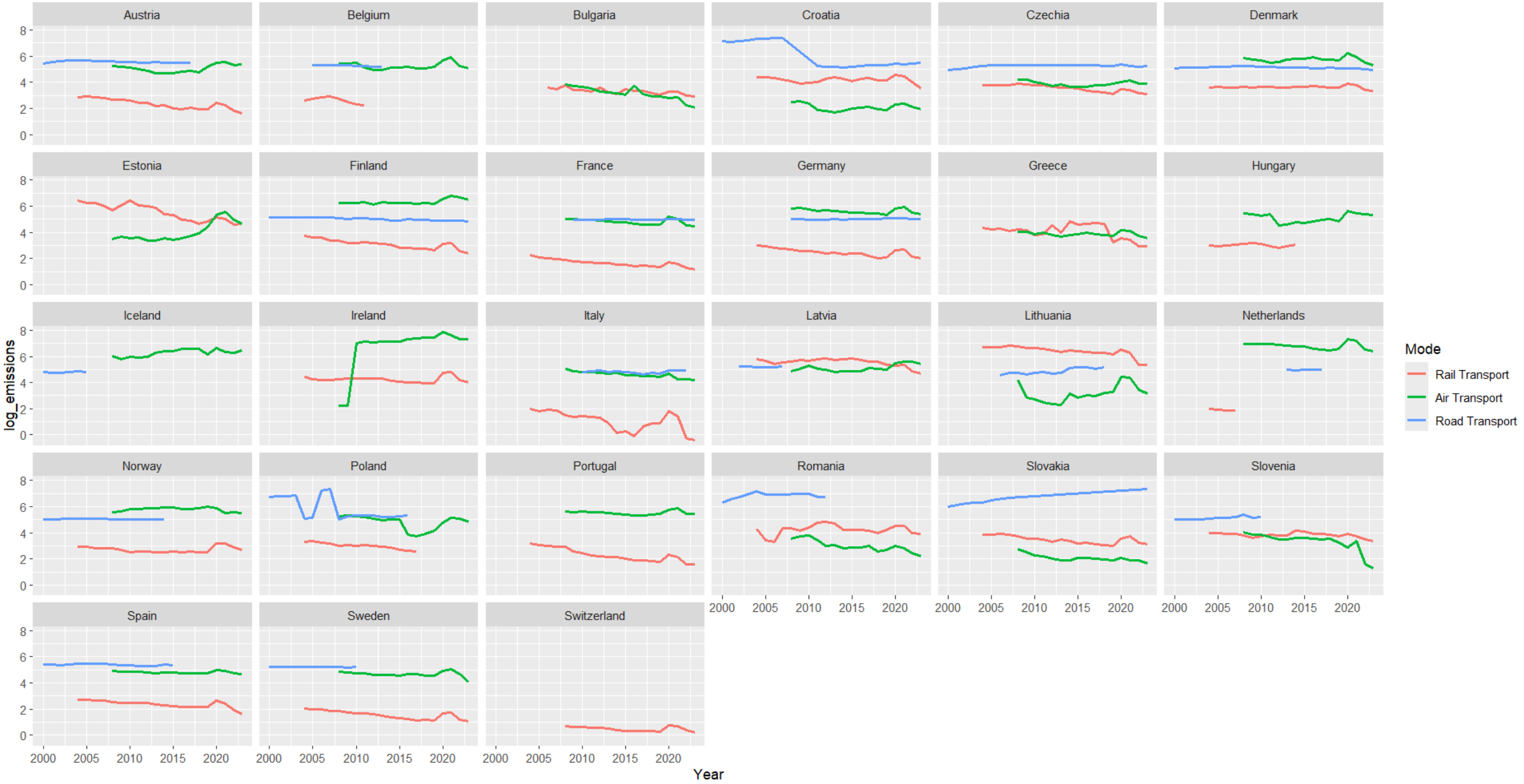
Table 1.4: Summary of single factor (GDP per capita) linear model for Air- Rail for the year 2023

Coefficient estimate	Std. error	p-value
(Intercept) = -736.0	215.6	0.00275
(GDP) = 0.02982	0.005018	$8.21 * 10^{-6}$

Emissions: Modeling

- Random intercept model underwent several trials
  - Log transform emissions to stabilize variance
  - Removed Malta and Luxembourg from dataset (small countries < 1,000 mi2)
  - Introduced new predictor land area; no significant effect
  - Centered GDP to address collinearity with intercept

Emissions: Modeling



Emissions: Modeling

- Final model (random intercept)

  $\log \text{emissions}_{ij} = \beta_0 + \beta_1 I_{\text{Air}} + \beta_2 I_{\text{Road}} + \beta_3 \text{gdp}_{ij} + u_j + \epsilon_{ij}$,

  where $i = \text{observation (year)}$ and $j = \text{country}$
- Fixed effect estimates

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	3.31707	0.17054	26.25711	19.450	< 2e-16
Air Transport	1.25992	0.07455	1151.89126	16.901	< 2e-16
Road Transport	2.03416	0.08729	1159.60178	23.303	< 2e-16
log(gdp_centered)	-0.54903	0.18310	98.48988	-2.999	0.00343

- The reference mode is rail, and we see that the model predicts air and road are less efficient

  Given GDP, Air is $e^{1.26} \approx 3.5$ times less efficient, and road is $e^{2.03} \approx 7.6$ times less efficient (per pkm)

  Per log-unit increase above average GDP per capita, a country emits $e^{-0.55} \approx 0.58 \times$ as much CO_2/PKM

Emissions: Modeling

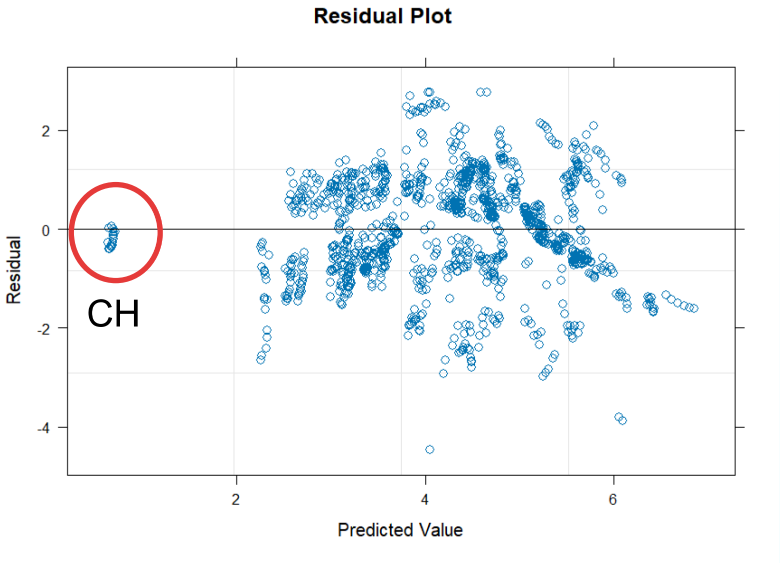
- Random effects
  - Most countries have random intercept effects less than 1, and Iceland, Ireland, and the Netherlands are the only countries with effects greater than 1
  - Switzerland has random effect of about -2

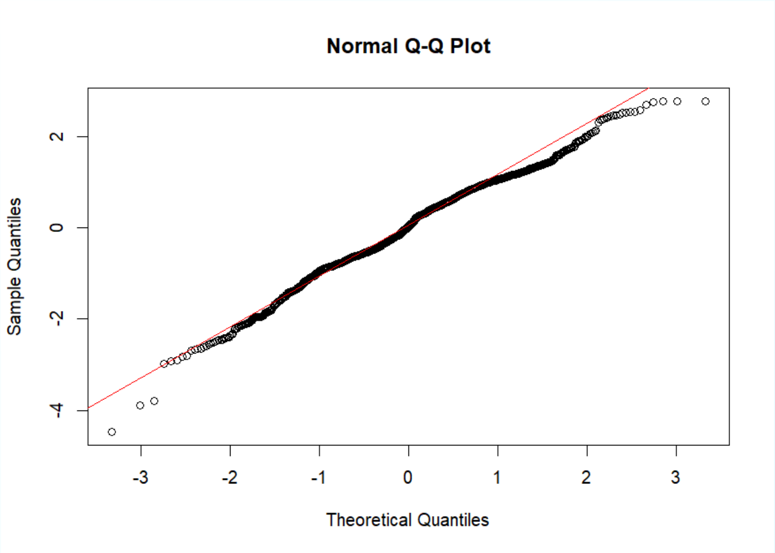
Variance estimates:

Groups	Name	Variance	Std.Dev.
Country	(Intercept)	0.7375	0.8588
Residual		1.1666	1.0801

Emissions: Modeling

- Diagnostics





Emissions: Extrapolation

- Extrapolation to 4 U.S. States
  - Used predInterval() from merTools package
  - Compute 80% prediction intervals
  - These intervals only consider fixed effects
  - Use GDP data from 2023

Texas

- Rail

  Estimate: 28.27, Interval: (7.44, 122.46)
- Air

  Estimate: 101.54
- Road

  Estimate: 224.25

Emissions: Extrapolation

California

- Rail
  Estimate: 26.69, Interval: (6.72, 111.01)
- Air
  Estimate: 89.59
- Road
  Estimate: 215.29

Colorado

- Rail
  Estimate: 27.16, Interval: (6.63, 110.93)
- Air
  Estimate: 98.10
- Road
  Estimate: 226.38

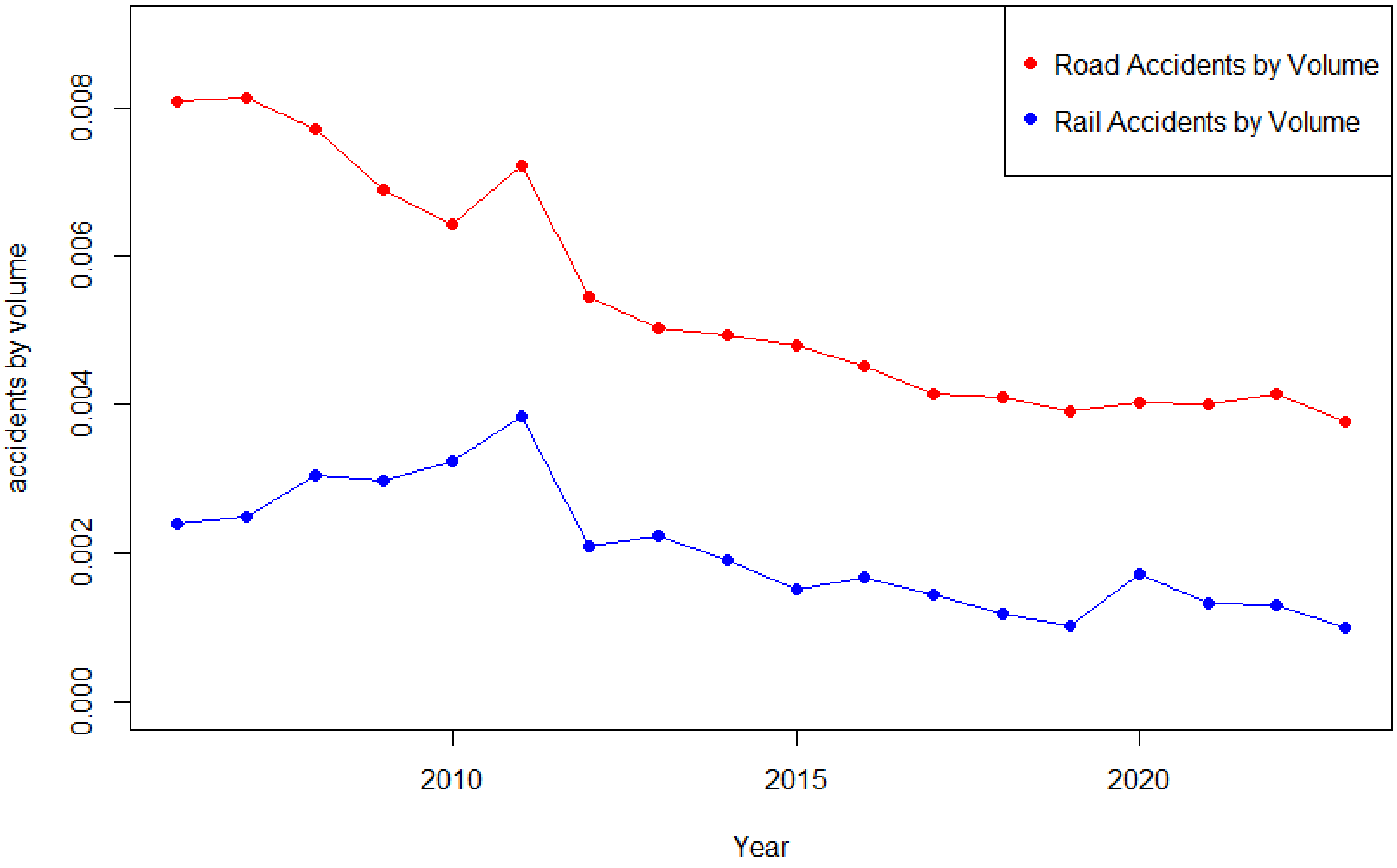
Minnesota

- Rail
  Estimate: 29.75, Interval: (7.27, 112.67)
- Air
  Estimate: 106.51
- Road
  Estimate: 216.98

New York

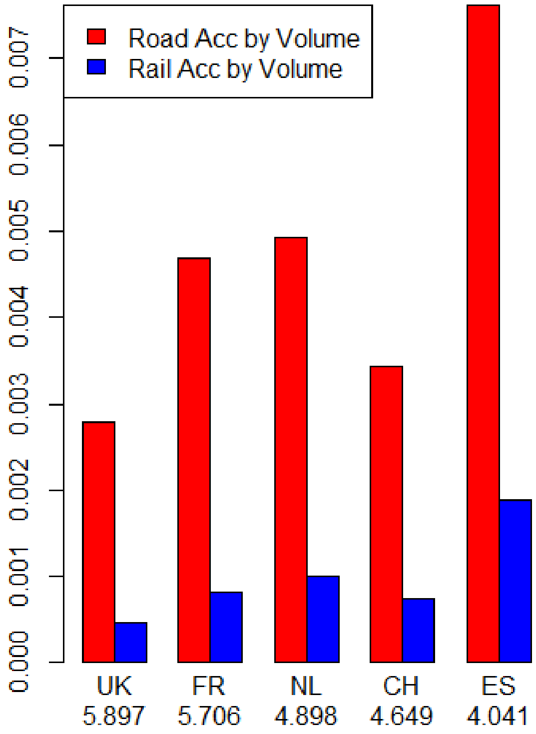
- Rail
  Estimate: 24.54, Interval: (6.15, 99.61)
- Air
  Estimate: 91.27
- Road
  Estimate: 197.50

03 Safety



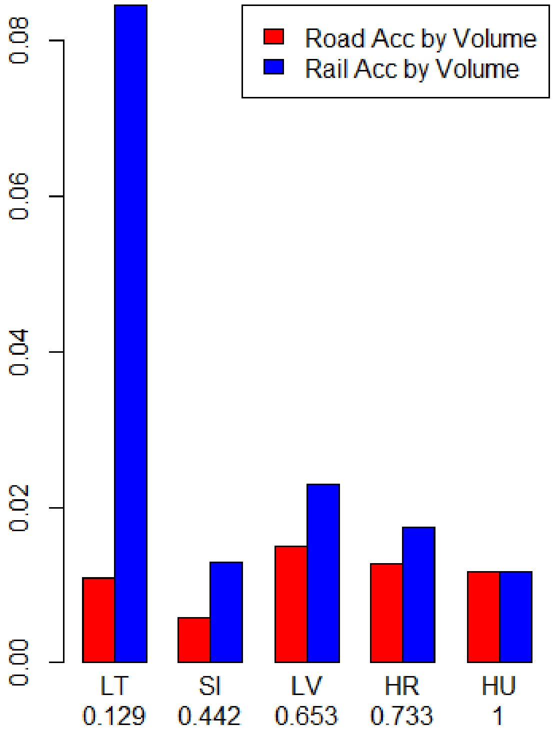
Road Accidents by Volume
Rail Accidents by Volume
accidents by volume
0.000
0.002
0.004
0.006
0.008
2010
2015
2020
Year

Fig 2.2



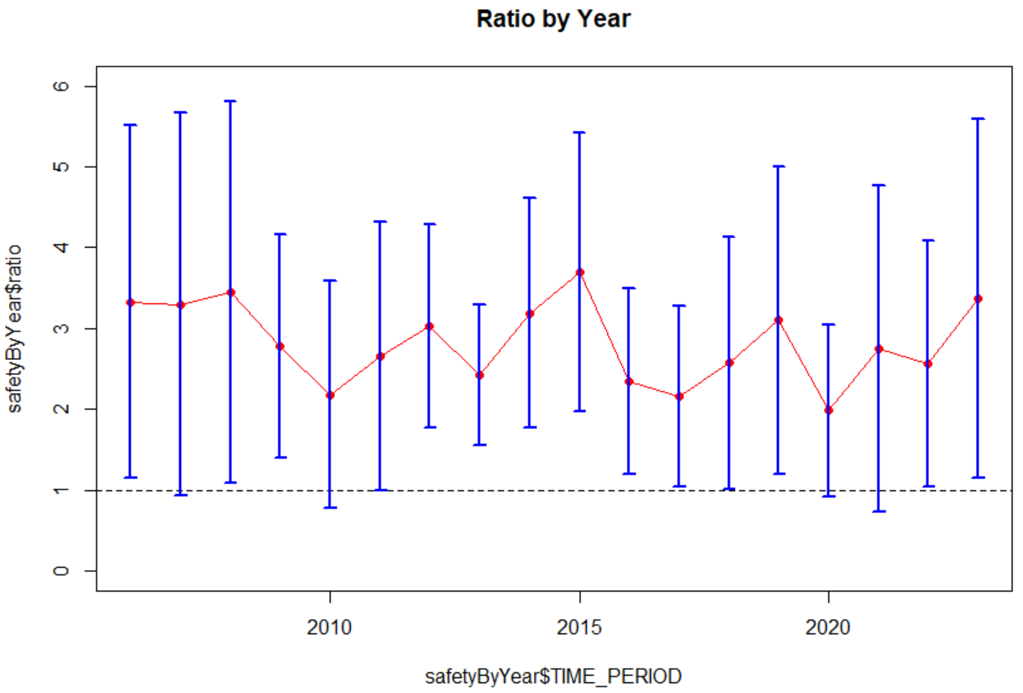
Road Acc by Volume
Rail Acc by Volume
0.000
0.001
0.002
0.003
0.004
0.005
0.006
0.007
UK
5.897
FR
5.706
NL
4.898
CH
4.649
ES
4.041

Fig 2.3



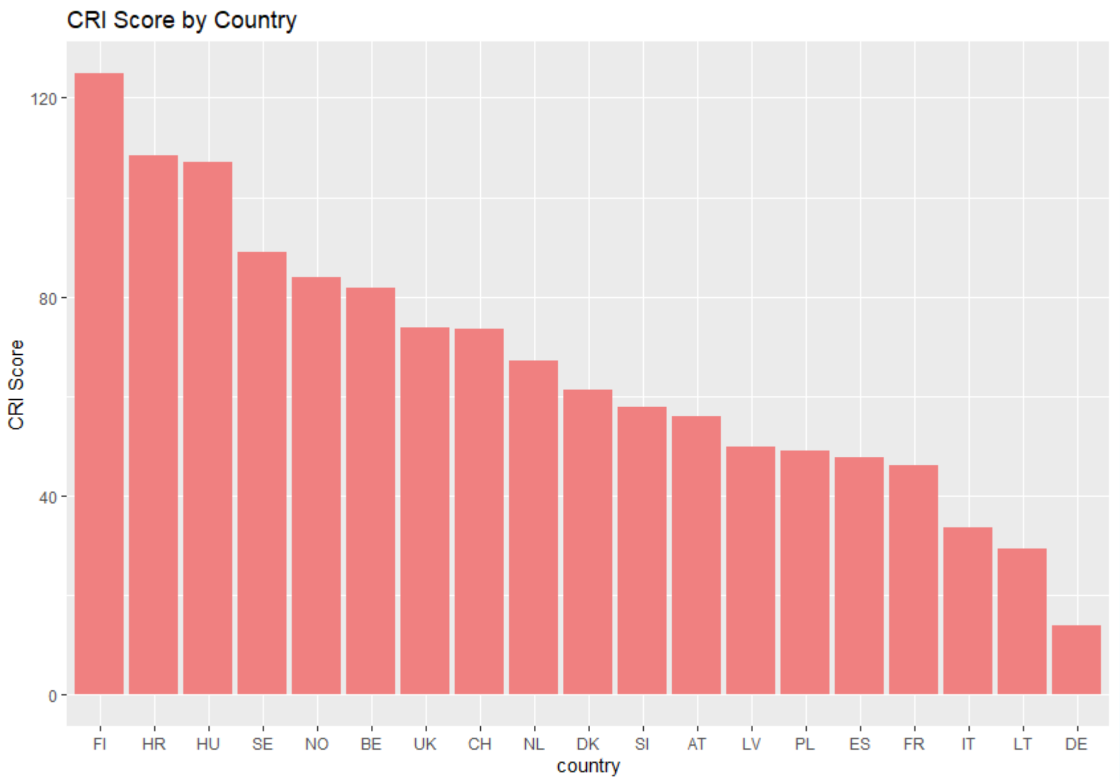
Road Acc by Volume
Rail Acc by Volume
0.00
0.02
0.04
0.06
0.08
LT
0.129
SI
0.442
LV
0.653
HR
0.733
HU
1

Hypothesis Testing



- 1-Sided Paired T-test across countries from 2006-2023
- H0: μ = 1 Ha: μ > 1
- 95% 2-Sided Confidence Intervals for each year
- 2009-10 – Dip due to recession
- 2014-15 – Large decrease in Rail Accidents by Volume in Spain, France and the UK due to new infrastructure
- 2020 – Pandemic cause a lot of rail shutdowns causing lower passenger volumes while the rail accidents didn't change that much

Modeling: EDA

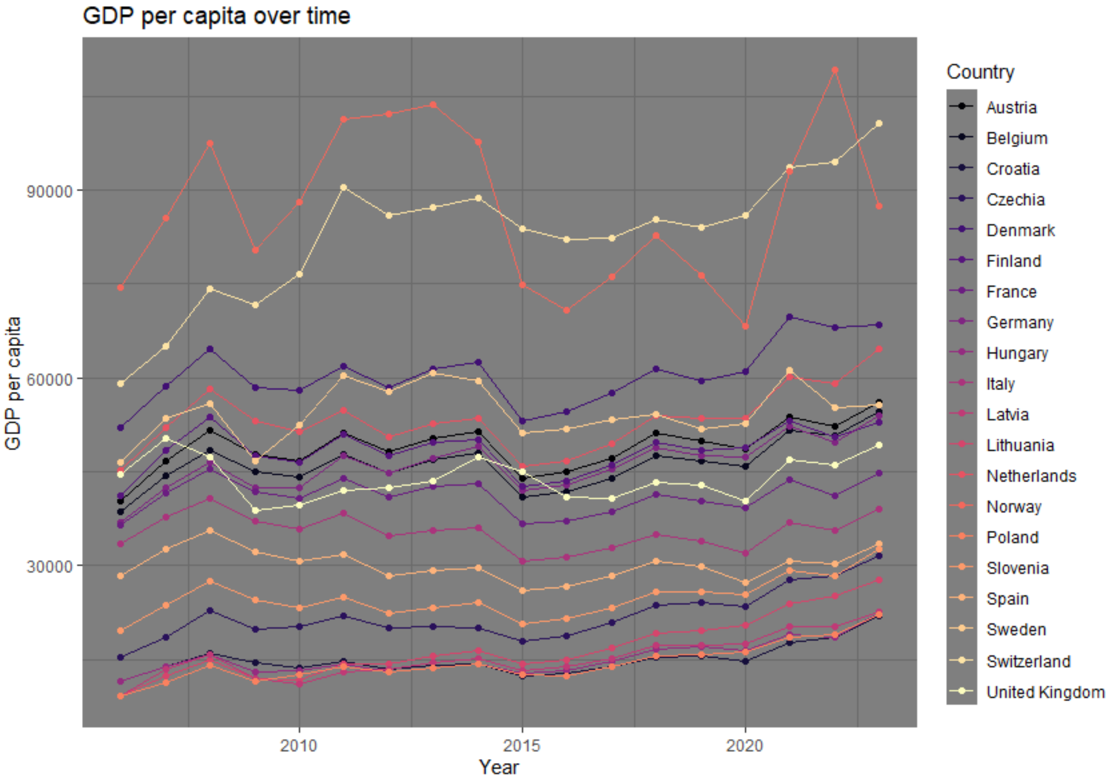


- High Scores: Finland, Sweden, Norway, Hungary, Croatia
- FI, SE, NO – quite northern
- HU, HR – quite mountainous

- Low Scores: Germany, Lithuania, Italy, France, Spain
- Germany, Italy, France, Spain – Large countries with lots of flat land
- Lithuania – Incredibly flat topography

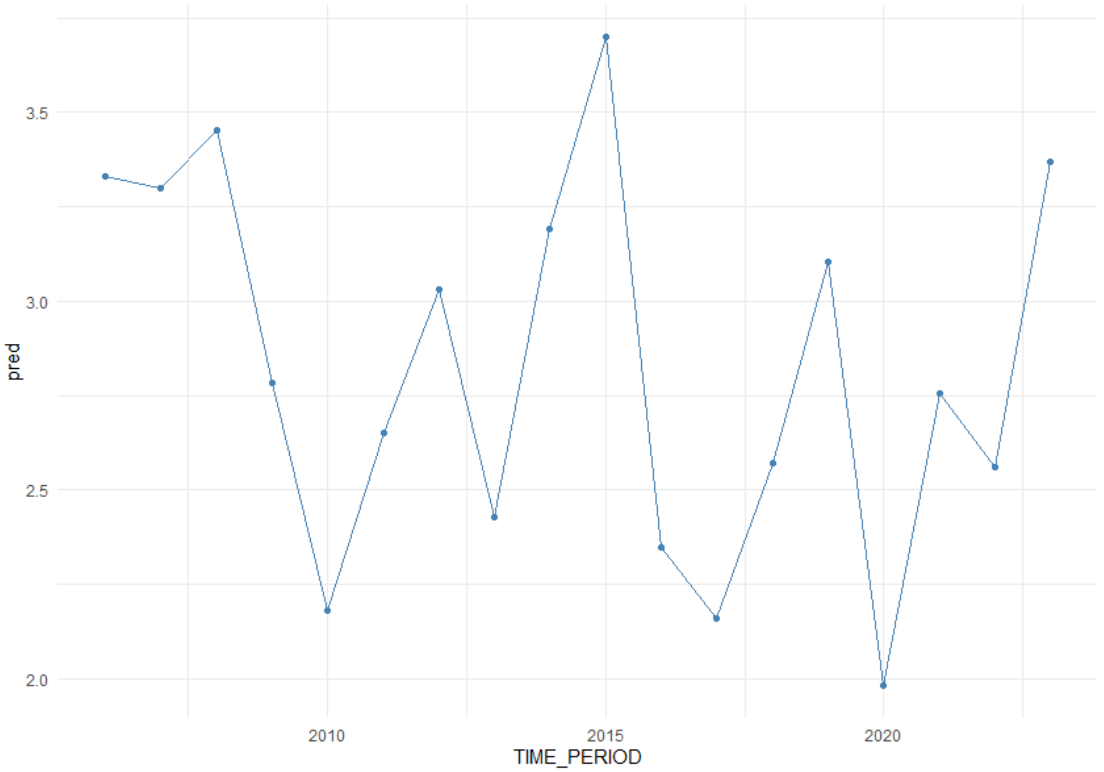
- Trains struggle with elevation changes as well heavy weather effects

Modeling: EDA



- High: Switzerland, Denmark, and Norway
- Relatively small populations with high amounts of wealth. Known for higher social infrastructure.
- Middle: Germany, UK, France, Spain
- High populations and accompanying wealth match each other.
- Bottom: Slovenia, Croatia, and Latvia
- Younger countries without much industry

Modeling: Linear by Year



- Identified trends in significant years
- Only one year with non-normal residuals. The UK in 2016 was an outlier due to the beginning of railway strikes causing the ratio to spike.
- CRI Score – 1/18
- Negative so more weather events means roads were safer
- GDP – 11/18
- Positive because more money means more money to put into infrastructure. Reflected in 2015
- Area – 7/18
- Positive because the bigger a country is, the higher number of urban areas develop. This causes increasing amounts of Road Accidents by volume due to increasing intra-city roads.

Modeling: Mixed Effects

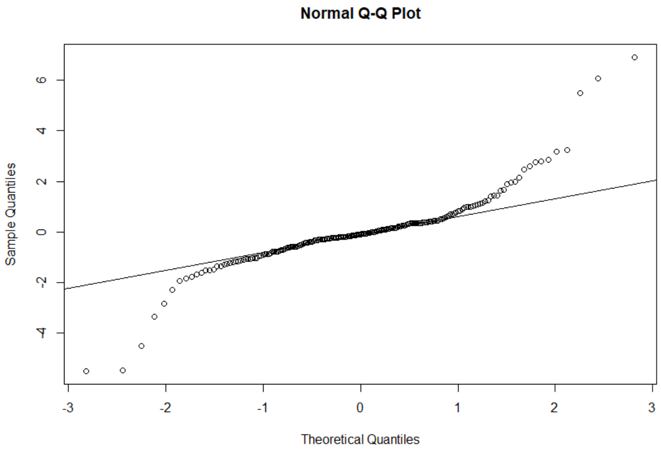
Source	Estimate	SE	DF	T	p
(Intercept)	-22.03	5.62	19.31	-3.92	0.0009
log(area)	0.63	0.30	14.05	2.08	0.056
log(gdp)	1.69	0.46	25.96	3.64	0.001

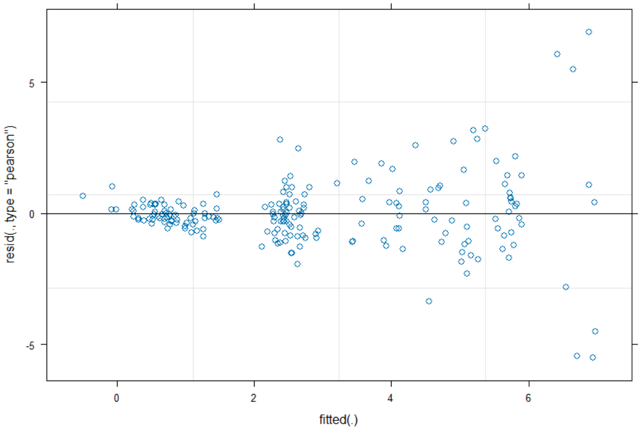
- GDP – very significant
- Area- only significant at 90% but kept in model for prediction accuracy
- CRI – removed as it was never significant ($p>0.25$) across all tested models

Groups	Variance	SD	% Var
geo	1.741	1.319	83.67%
Residual	2.067	1.438	100%

geo accounts for much of the variance due to differing intangibles between cultures that are not accounted for in the model

Modeling: Mixed Effects GoF





- AIC = 777.22
- BIC = 793.8106
- $R^2_{conditional}$ = 0.6342995
- $R^2_{marginal}$ = 0.328275
- Extremely heavy tails though no signs of skewness
- Heteroskedasticity
- As countries have higher GDP per capita that is usually correlated with a greater amount of differing lifestyles
- Bigger countries have greater space for locations with differing transport cultures

Extrapolation

State	EU	Marginal	Marginal CI	Conditional	Conditional CI
TX	NO	5.53	(2.330,8.58)	7.07	(3.75,10.12)
CA	CH	5.70	(2.40,8.58)	6.92	(3.65,9.99)
CO	NO	5.20	(2.24,8.3)	6.54	(3.36,9.72)
MN	UK	4.78	(1.79,7.92)	6.79	(3.63,10.04)
NY	CH	5.18	(1.86,8.14)	6.52	(3.45,9.59)

Thanks!

Questions?

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution

Sources

"Breakeven Distances between Conventional Rail, High Speed Rail and Air Transportation." The Spatial Organization of Transportation and Mobility, The Geography of Transport Systems, 18 Apr. 2024, transportgeography.org/contents/applications/high-speed-rail-systems/breakeven-distances-rail-air-transport/.

Bureau of Transportation Statistics. United States Department of Transportation, bts.gov.

Eurostat. "Air emissions accounts by NACE Rev. 2 activity (env_ac_ainah_r2)." Eurostat, ec.europa.eu/eurostat/databrowser/view/env_ac_ainah_r2__custom_18136650/default/table.

Eurostat. "Database – Transport." *Eurostat*, ec.europa.eu/eurostat/web/transport/database.

European Environment Agency. "Air quality statistics." *European Environment Agency*, 31 Oct. 2025, eea.europa.eu/en/analysis/maps-and-charts/air-quality-statistics-dashboards?activeTab=570bee2d-1316-48cf-adde-4b640f92119b.

U.S. Census Bureau. "Annual Estimates of the Resident Population for the United States, Regions, States, District of Columbia, and Puerto Rico: April 1, 2020 to July 1, 2024 (NST-EST2024-POP)." State Population Totals: 2020-2024, www2.census.gov/programs-surveys/popest/tables/2020-2024/state/totals/NST-EST2024-POP.xlsx.

Eurostat. "Railway transport." *Eurostat*, ec.europa.eu/eurostat/web/transport/information-data/railway-transport.

World Bank. "Land area (sq. km) (AG.LND.TOTL.K2)." World Development Indicators, databank.worldbank.org/reports.aspx?source=2&series=AG.LND.TOTL.K2&country=EUU.