Preprint:
Bridgelall, R., and Tolliver, D. (2023). Quantifying Freight Flow Disruption Risks from Railroad Accidents. *Quality & Quantity*, August 28, 2023. DOI: 10.1007/s11135-023-01727-3.

Quantifying Freight Flow Disruption Risks from Railroad Accidents

Raj Bridgelall, Ph.D. (Corresponding Author)
Associate Professor
Transportation, Logistics & Finance, College of Business
North Dakota State University
PO Box 6050, Fargo ND 58108-6050
Email: raj@bridgelall.com, ORCID: 0000-0003-3743-6652

Denver D. Tolliver, Ph.D.
Director, Upper Great Plains Transportation Institute
North Dakota State University
PO Box 6050, Fargo, ND 58108
Email: denver.tolliver@ndsu.edu, ORCID: 0000-0002-8522-9394

Statements and Declarations

Funding:
The authors conducted this work with support from North Dakota State University and the Mountain-Plains Consortium, a University Transportation Center funded by the U.S. Department of Transportation.

Competing Interests: The authors have no relevant financial or non-financial interests to disclose.

Author Contributions:
R.B.: conceptualization, methodology, software, data curation, formal analysis, writing—original draft preparation. D.T.: supervision, resources, funding acquisition, project administration, validation, writing—reviewing and editing.

Data Availability:
The data that support the findings of this study are openly available at the sources cited within the manuscript.

Quantifying Freight Flow Disruption Risks from Railroad Accidents

Abstract

Workforce shortages during the COVID-19 pandemic and the recent threat of railroad strike in the United States have generated greater public awareness of how freight flow disruptions can harm society. Unlike highways, trains often do not have the ability to take alternative routes that directly connect major metropolitan areas; hence railroad networks are vulnerable to disruptions like accidents. This study developed a data mining workflow to rank commodity movements that are at the highest risk of disruption from railroad accidents and other types of regional disasters that can affect railroad operations. A key finding is that five U.S. metropolitan areas are at least five times more likely than others to experience a railroad accident. Those five areas account for more than 40% of the monetary value in alcoholic beverages, raw meats, gasoline, plastic-based products, and rubber-based products moved by rail. Hence, any disruption in those five areas can lead to widespread shortages of those commodities. The implication is that decision makers should focus risk mitigation and resiliency strategies in those five metropolitan areas and on the top commodity categories at risk.

Keywords: Capacity disruption; Commodity flows; Risk assessment; Risk management; Supply chain resilience; Sustainable supply chains

1 Introduction

The recent threat of railroad strike in the United States increased public awareness of how important railroads are to the economic and social well-being of the nation. Railroad commodity flows are vulnerable to any sort of blockage on a line because the networks lack direct alternative routes between major metropolitan areas. Weather-related disasters such as heavy snowfall, flooding, or falling debris can block traffic (Sharma, et al. 2021). Accidents can block or damage shipments, thereby creating severe commodity shortages. The loss of railroad capacity at seaports and other intermodal transloading facilities can result in backlogs and spoilages as vessels queue for processing (Schofer, Mahmassani and Ng 2022). In a multimodal or intermodal network, a railroad accident can affect freight flows in other modes that use highways, waterways, or airways (Kelle, et al. 2019). The loss of freight from accidents or theft can have cascading effects because manufacturers rely on the timely delivery of raw materials to produce finished goods. Disruptions of even a single commodity flow can affect multiple industries that rely on that item for economic productivity (Darayi, Barker and Nicholson 2019).

The risk of railroad accidents in North America is predictable because the incidents have consistently hovered around 2,500 per year (Bridgelall and Tolliver 2022). The dominant causes of railroad accidents are consistently human error and problems with the track or roadbed (Bhardwaj, et al. 2021). Positive train control (PTC) systems installed in North America to reduce the risk of human errors make railroads more connected than ever (Zhang, Liu and Holt 2018). Therefore, new cyber-threats to railroads are likely to develop, which can disable key railroad systems and operations (Kolli, Lilly and Wijesekera 2018).

Greater awareness of the potential disruption impact from railroad accidents will inform risk mitigation strategies. Therefore, the goal of this research is to quantify the risk of railroad accidents among the metropolitan statistical areas (MSAs) of the continental United States (CONUS). The approach identified MSAs that are accident outliers and then ranked the types and proportion of commodities that they moved by rail. The results of this work will help create awareness and insights for stakeholders to prioritize and tailor risk mitigation strategies and resource planning to enable a more resilient system. The data mining framework presented is applicable to any region of the world.

The United States Census Bureau (USCB) defined 82 MSAs of the CONUS based on their population and importance as transportation hubs or foreign trade gateways (BTS and USCB 2020). Using a Geometric Information System (GIS) tool and shapefile data from the USCB to produce Figure 1, the relative size and spatial distribution of the MSAs on the CONUS becomes evident. The figure also shows the arrangement of the MSAs relative to the major railroad tracks that serve them. This visualization illustrates how the present railroad network will require significant diversions to find alternative routes between MSAs. Consequently, rerouting will cause delivery delays and increased transport costs until agencies can complete their investigations and clear the incident to reopen railroad tracks.

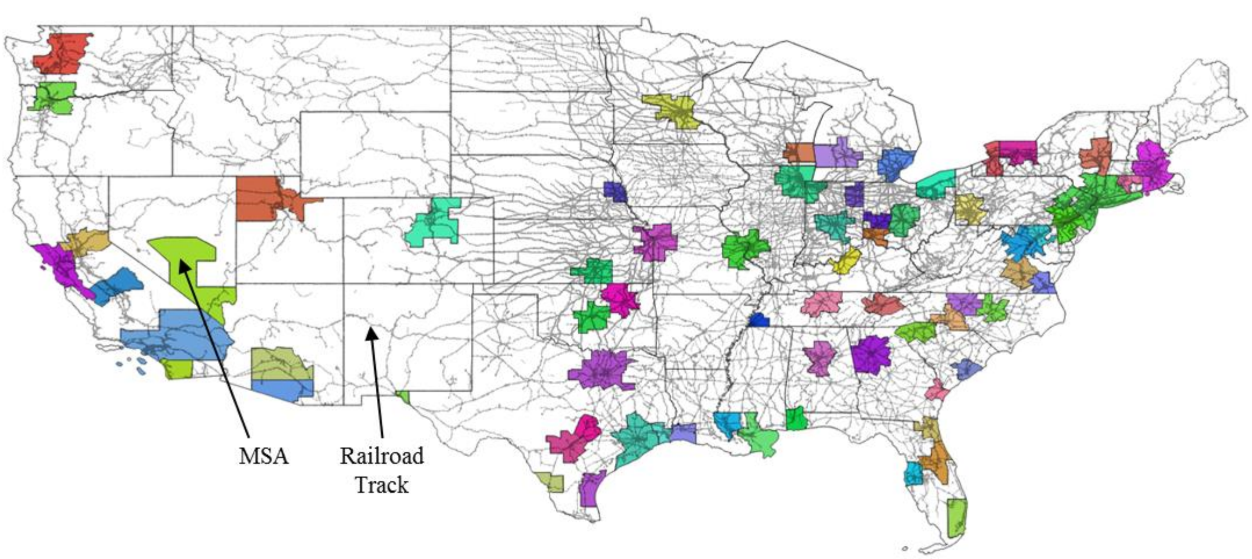

Figure 1: MSAs of the CONUS and the major railroad routes that serve them.

Recent media attention to railroad accidents in the United States has raised awareness about their implications to society. A few studies examined on how extreme weather events can delay railroad operations (Huang, et al. 2020) and even cause widespread disruptions to the supply chain (Woodburn 2019). Analysts are increasingly interested in understanding factors that can affect the resilience of rail transport (Bešinović 2020). To gain insights, more researchers are beginning to apply data mining to large datasets on operations, safety, and maintenance (Ghofrani, et al. 2018). Especially vulnerable to point disruptions are freight bottlenecks at intermodal terminals such as ports, which are the main gateways of foreign trade, inland terminals, and transshipment yards (Basallo-Triana, Bravo-Bastidas and Vidal-Holguín 2022). A striking example is how human resources shortages during the COVID-19 pandemic exacerbated delays at seaports (Schofer, Mahmassani and Ng 2022).

Jabbarzadeh et al. (2020) found that contingency plans can significantly mitigate the risk of disrupted hazardous material transport by rail for only a slight increase in total cost (Jabbarzadeh, Azad and Verma 2020). Procházka et al. (2020) conducted a statistical evaluation of accidents involving the transport of dangerous substances and found that broad implementation of safety regulations reduced the impact on affected populations (Procházka, Hošková-Mayerová and Procházková 2020). Overall, there has been few academic studies about how the disruption of railroad capacity can affect trade bottlenecks (Wendler-Bosco and Nicholson 2020).

The organization of the rest of this paper is as follows: Section 2 describes the methodology to identify accident outlier MSAs and the commodity flows that will be most affected. Section 3 discusses the results. Section 4 discusses the implications for society and potential solutions. Section 5 concludes the paper and suggests future work.

2 Methodology

The next subsections describe the datasets, the technique used to combine them, the workflow to identify high-risk metropolitan areas, and the commodity flows at highest risk.

2.1 Data Sourcing

Figure 2 illustrates the analytical workflow to rank the risks of commodity flow disruption from railroad accidents. The input datasets were 1) the Rail Equipment Accident database from the Federal Railroad Administration (FRA) (FRA 2021), 2) the freight analysis framework (FAF) data from the Federal Highway Administration (FHWA) (FHWA 2021), and 3) the commodity flow survey (CFS) geography definitions (USCB 2021). The FRA accident database contains data from the mandatory reporting of accidents that resulted in a certain amount of damage cost. The cost threshold varies annually—it was $11,300 for the year 2022 (FRA 2021). In addition to dozens of fixed fields that describe features of each accident, such as location and damage amount, the reports also contain 15 narrative fields that provide further details about the event (Bridgelall and Tolliver 2021).

The CFS is a joint effort by the Bureau of Transportation Statistics (BTS) of the United States Department of Transportation and the United States Census Bureau (USCB) of the Department of Commerce (BTS and USCB 2020).

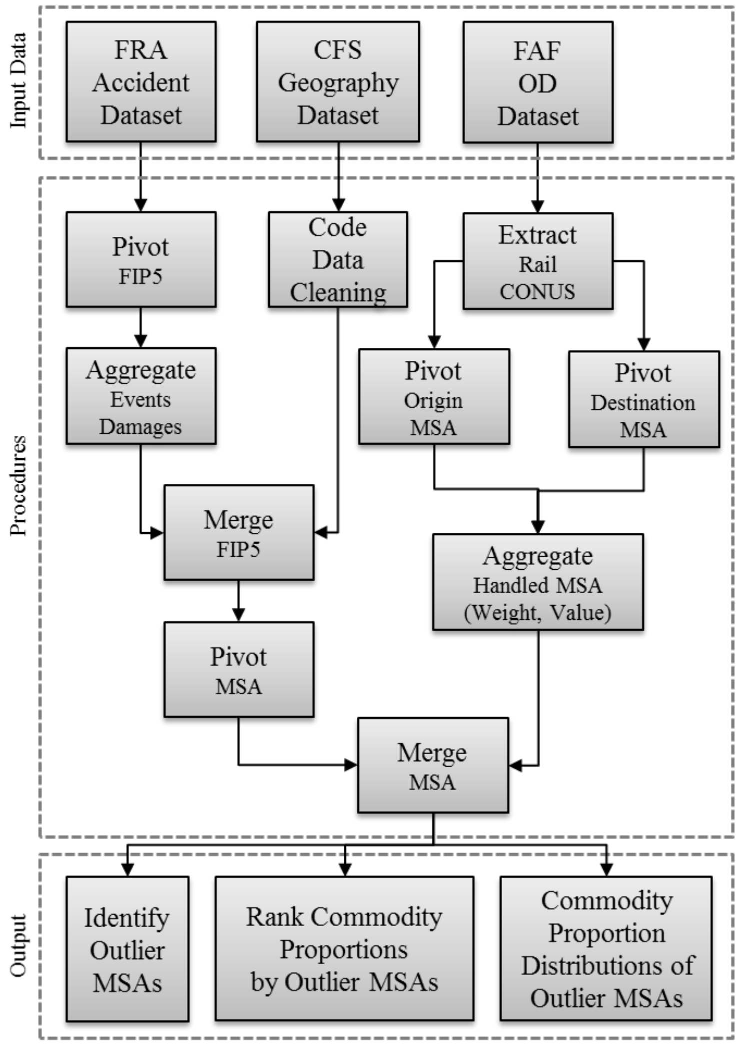

Figure 2: Input datasets and analytical workflow.

The two agencies conduct the survey every five years as part of the Economic Census. The CFS is the primary source of domestic freight shipment data of the CONUS. The CFS primarily covers shipping data from the mining, manufacturing, and wholesale sectors. The FAF dataset, which is a joint production of the BTS and the FHWA, extends the CFS data (FHWA 2021). The FAF dataset integrates shipping data from additional sectors such as agriculture, extraction, utility, construction, and service. The FAF version 5.2 dataset, published in 2021, provides estimates for commodities moved between MSAs in 2017 by weight (thousand tons) and by value (million dollars). The movements by rail include common or private railroad carriers but excludes rail movements that are part of multiple modes and mail.

2.2 Data Merge

The workflow extracted railroad accidents that occurred during the decade from 2009 through 2019 to reflect railroad movements prior to disruptions from the COVID-19 pandemic. There were 28,237 rows of data remaining after extracting records for the CONUS. Each row contained 145 fields of information relating to a railroad accident. The subset of features used for the analytical workflow were the date, accident damage value in U.S. dollars, and the federal information processing (FIP) code for the county where the accident occurred. The CFS geography dataset contained the names, FIP codes, CFS area codes (CFS07_DDESTGEO), and the 2017 CFS area names (CFS17_NAME) associated with 3,143 U.S. counties. The 132 CFS area codes matched the FAF region codes, except for 13 MSAs. Consequently, the workflow incorporated data cleaning to identify and substitute the mismatched codes as summarized in Table 1.

Table 1: Cleaning of CFS Geography Dataset

CFS Area Name	CFS Code Replaced	FAF Code Substituted
Fresno-Madera, CA CFS Area	69	65
Philadelphia-Reading-Camden, PA-NJ-DE-MD (DE Part)	100	101
Remainder of Delaware	100	109
Fort Wayne-Huntington-Auburn, IN	189	183
Wichita-Arkansas City-Winfield, KS	209	202
Louisville/Jefferson County-Elizabethtown-Madison, KY-IN (KY Part)	211	212
Omaha-Council Bluffs-Fremont, NE-IA (NE Part)	310	311
Remainder of Nebraska	310	319
Boston-Worcester-Providence, MA-RI-NH-CT (NH Part)	330	331
Remainder of New Hampshire	330	339
New York-Newark, NY-NJ-CT-PA (PA Part)	429	423
Knoxville-Morristown-Sevierville, TN	479	473
Portland-Vancouver-Salem, OR-WA (WA Part)	539	532

A string processing procedure then created a data merge key by concatenating the two-digit FIP code for the state and the three-digit FIP code for the county to form a unique five-digit county code (FIP5) that matched the FIP code in the FRA dataset. The workflow then merged the FRA accident and CFS geography datasets by the FIP5 codes. Subsequently, a pivot table procedure grouped and aggregated the number of accidents and accident damage costs by the FAF region codes.

Rows of the FAF 5.3 dataset contained the tonnage and value of a commodity class by their origin and destination FAF region codes and the transportation mode. Extracting the CONUS rail-only movements reduced the dataset from 1,627,492 rows to 175,816 rows. Pivot table procedures then aggregated the weight and value of commodity categories by FAF region origin and destination codes. The procedure then aggregated the commodity categories leaving (origin) and entering (destination) by weight and value handled within each FAF region. A merge procedure then combined the summarized FRA and FAF datasets by the FAF region codes. Of the 132 FAF regions, 82 were MSAs on the CONUS and the others were the remainder of a state. The final procedure retained the data for the MSAs.

2.3 Outlier Identification

The method produced a histogram (Figure 3 in the results section) of the decade accumulated accident counts among MSAs to identify the outliers. An optimization procedure fitted a lognormal distribution to the histogram without outliers and evaluated the goodness of fit with a chi-squared statistical test. The optimization procedure was as follows:

$$\underset{X_i}{\text{minimize}} \quad e = \sum_{i=1}^{B} (H_i - f_i)^2$$
$$\text{subject to} \quad \alpha > 0, \sigma > 0, \mu > 0 \text{ and } N \geq B \geq 4 \tag{1}$$
$$\text{where} \qquad f_i = \frac{\alpha}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}}, \; i = 1, 2, \ldots, B$$

The parameters α, σ, and μ are the amplitude, standard deviation, and mean, respectively, of the distribution function, f. The starting interval for histogram bin i was X_i and the number of MSAs.

The number of accidents that fell within that interval was H_i. The distribution under test was the function f_i. The solution to the optimization problem was the combination of amplitude α, mean μ, and variance σ^2 of f_i that minimized the sum-of-squares error e, subject to the constraints indicated. The test statistic was the Pearson's chi-squared statistic of k degrees of freedom (*df*) where

$$\chi_k^2 = \sum_{i=1}^{B} \frac{(H_i - D_i)^2}{D_i} \tag{2}$$

and the p-value associated with the statistic guided whether to reject the null hypothesis that the histogram followed the tested distribution. The constraint on the minimum number of bins B was at least 4 because with three parameters (α, μ, and σ) estimated, the minimum *df* was unity (4 - 3 = 1). Outlier MSAs are those with accumulated accident counts greater than 1.5 standard deviation above the mean.

3 Results

Figure 3 shows the histogram of railroad accidents accumulated among MSAs during the decade ending in 2019. The minimum and maximum number of accumulated accidents among MSAs was 5 and 1,550, respectively. The mean was 183.6 and the standard deviation was 231.1. Therefore, the outlier threshold was 530.3. The five callouts show that the MSA accident outliers were within and around the cities of Chicago (IL), Houston (TX), Los Angeles or LA (CA), Dallas-Fort Worth or DFW (TX), and Newark (NJ). The chi-squared statistic with *df* of 28 was 7.85 and the p-value was nearly 1.0. The customary p-value threshold for hypothesis rejection is 0.05 (Agresti 2018). Therefore, the test could not reject the null hypothesis that the distribution followed the lognormal.

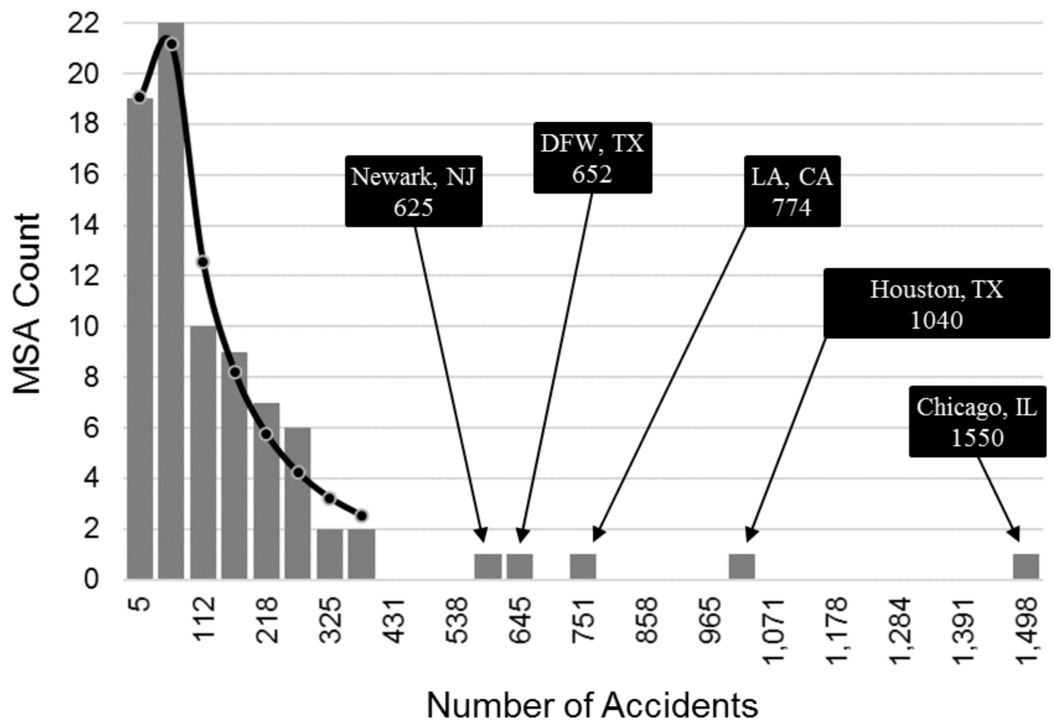

Figure 3: Distribution of the number of railroad accidents accumulated from 2009 to 2019.

The first four columns of Table 2 summarize the sum and mean values for accident count, accident damage amount, commodity weight moved, and United States dollar (USD) value of commodity moved in all MSAs.

Table 2: Accident and Commodity Statistics of Outlier MSA Proportions

Feature	Units	Statistic	MSAs	Outliers	Outliers %
Accidents	Count	Sum	15057	4641	30.8%
Accidents	Count	Mean	183.6	928.20	505.5%
Damage	USD	Sum	$1,701,398,934	$479,271,757	28.2%
Damage	USD	Mean	$13,189,139	$95,854,351	726.8%
Weight	Thousand Tons	Sum	1,377,862.24	233,014.24	16.9%
Weight	Thousand Tons	Mean	10,681.10	46,602.85	436.3%
Value	Million USD	Sum	739,870.42	162,706.37	22.0%
Value	Million USD	Mean	5,735.43	32,541.27	567.4%

The “Outliers” and “Outliers %” columns summarize the same statistics for the five outlier MSAs and the proportion they accounted for, respectively. For example, the first row indicates that sum of accidents within the outlier MSAs accounted for 30.8% of the railroad accidents in all MSAs. The last row indicates that the mean USD of commodities moved by the five outlier MSAs was more than 567% greater than that moved by the average MSA. The ranking compares accident risk among MSAs where freight flows originate or terminate, hence the workflow does not include accidents that occurred outside of MSAs. Table 3 lists the weight proportion of each commodity category moved by each of the outlier MSAs. The table ranks the commodity categories by the sum of their weight proportions across the outlier MSAs, starting with the highest proportion. Table 4 provides a similar list as Table 3, but instead shows the monetary value proportion of the commodity categories moved. The last column of Table 3 shows that the five outlier MSAs moved more than 61% of the weight of alcoholic beverages transported by rail in the CONUS. The BTS lists all commodities associated with the assigned standard classification of transported goods (SCTG) codes listed in the table (BTS and USCB 2015). The last column of Table 4 shows that the five outlier MSAs moved more than 53% of the USD value of raw meats transported by rail in the CONUS.

Table 5 lists the weight and monetary value proportion of the commodity categories moved by the top outlier MSAs. The table rows are in the order of the commodity monetary value proportion, starting with the highest proportion. As observed, alcoholic beverages ranked as the top commodity moved by rail with the Chicago MSA accounting for more than 61% and 56% of its handled weight and monetary value proportions, respectively. The top five accident outlier MSAs moved 19 of the 42 commodity categories at risk. Los Angeles was the top MSA for eight of the commodity categories moved, namely raw meats, printed products, furniture, electronics, products of animals, prepared foods, miscellaneous manufactured goods, textiles, and leathers. Houston was the top MSA for seven of the commodity categories moved, namely gasoline, plastic articles, rubber articles, basic chemicals, chemical products, fuel oils, base metal articles, and coal.

Table 3: Weight Proportion of the Commodity Categories Moved by the Outlier MSAs

Commodity	SCTG Code	Chicago IL	Houston TX	LA CA	DFW TX	Newark NJ	Weight %
Alcoholic Beverages	08	47.5	5.2	6.5	0.7	1.6	61.5
Paper-based Products	28	36.7	0.8	9.1	0.3	0.8	47.8
Gasoline Fuel	17	9.5	19.3	2.5	4.0	11.5	46.8
Raw Meats	05	2.9	3.1	30.9	1.2	2.4	40.6
Plastics & Rubbers	24	3.7	30.4	1.9	3.1	0.9	40.0
Base Metal Articles	33	4.0	23.0	10.6	1.0	0.5	39.1
Basic Chemicals	20	3.2	20.6	9.3	2.3	0.9	36.3
Printed Products	29	2.1	1.5	29.9	0.5	1.7	35.8
Chemical Products	23	5.1	21.4	3.2	2.0	1.9	33.7
Furniture	39	3.0	6.1	15.9	0.5	7.9	33.4
Mixed Goods	43	2.2	8.4	12.9	6.5	2.3	32.4
Fuel Oils	18	0.5	16.9	2.7	7.6	2.1	29.8
Electronics	35	2.8	6.9	9.1	6.8	3.2	28.8
Pharmaceuticals	21	0.8	22.1	2.1	0.4	3.0	28.3
Milled Grains	06	3.3	0.1	12.1	0.2	11.8	27.6
Products of Animals	04	10.5	0.9	12.6	1.3	0.1	25.4
Transport Equipment	37	14.5	2.3	1.1	3.6	0.2	21.8
Petroleum Products	19	0.2	20.3	0.7	0.1	0.0	21.3
Waste & Scrap	41	3.3	2.0	5.7	1.8	7.6	20.4
Misc. Manufactured	40	1.1	1.5	12.1	0.7	4.6	19.9
Prepared Foods	07	6.3	2.4	5.9	1.0	1.1	16.7
Textiles & Leathers	30	1.2	2.1	5.8	1.0	5.2	15.3
Machinery	34	2.1	4.3	3.1	4.0	1.7	15.2
Crude	16	6.9	0.0	2.8	0.0	5.3	15.0
Precision Instruments	38	4.2	1.1	4.1	0.4	4.4	14.2
Wood Products	26	2.9	3.4	2.9	2.9	1.9	14.0
Agricultural Products	03	6.6	2.4	3.6	0.9	0.2	13.8
Coal	15	0.3	4.6	1.5	0.6	6.7	13.6
Non-metallic Minerals	13	8.9	1.4	0.8	1.0	1.0	13.1
Building Stone	10	0.0	2.5	2.9	0.8	4.1	10.3
Base Metals	32	2.3	2.6	1.4	2.7	0.7	9.7
Cereal Grains	02	5.0	2.4	0.5	1.2	0.4	9.5
Natural Sands	11	8.1	0.1	0.4	0.5	0.0	9.2
Newsprints & Papers	27	2.8	1.0	3.0	1.3	1.1	9.2
Mineral Products	31	2.3	3.7	1.3	1.3	0.5	9.1
Fertilizers	22	5.2	2.2	0.3	0.6	0.5	8.9
Motorized Vehicles	36	0.6	1.1	5.5	0.8	0.4	8.5
Logs	25	3.6	0.3	0.7	0.2	0.1	4.9
Gravel & Crushed Stone	12	0.2	1.1	0.1	3.1	0.0	4.5
Metallic Ores	14	0.7	0.2	2.0	0.1	0.5	3.4
Live Animals & Fish	01	0.0	0.0	2.8	0.0	0.0	2.8
Tobacco Products	09	0.0	0.0	0.2	0.0	1.7	1.9

Figure 4 shows a value proportion histogram for all the commodity categories moved by each accident outlier MSA. These histograms present another view of the data mining results. The trends indicate that each of the accident outlier MSAs tend to move by rail a much larger proportion (outlier) of some commodity categories than others.

Table 4: Monetary Value Proportion of the Commodity Categories Moved by the Outlier MSAs

Commodity	**SCTG Code**	**Chicago IL**	**Houston TX**	**LA CA**	**DFW TX**	**Newark NJ**	**Value %**
Alcoholic Beverages	08	41.5	4.5	4.7	0.8	4.8	56.3
Raw Meats	05	2.9	2.8	42.8	1.7	3.3	53.4
Gasoline Fuel	17	8.9	20.6	2.6	4.0	11.2	47.3
Plastics & Rubbers	24	4.1	27.1	4.4	3.1	1.3	40.0
Basic Chemicals	20	3.8	23.2	6.3	2.3	2.4	38.1
Printed Products	29	2.4	0.8	32.5	0.2	2.1	38.0
Chemical Products	23	5.3	20.7	6.0	1.6	3.1	36.7
Fuel Oils	18	0.9	22.8	2.7	7.3	2.0	35.8
Furniture	39	3.6	5.7	17.8	0.6	7.8	35.4
Electronics	35	3.4	7.8	11.9	8.0	4.0	35.1
Products of Animals	04	10.4	2.2	19.8	2.0	0.6	35.0
Base Metal Articles	33	4.3	21.3	5.6	1.2	1.1	33.5
Milled Grains	06	5.5	0.4	10.9	0.2	12.8	29.8
Mixed Goods	43	1.8	6.7	13.5	4.6	1.9	28.6
Petroleum Products	19	0.3	24.9	2.6	0.0	0.3	28.3
Pharmaceuticals	21	1.9	6.0	2.7	0.9	16.5	28.1
Building Stone	10	0.0	6.0	4.8	1.9	14.6	27.3
Prepared Foods	07	7.4	3.3	12.6	1.3	2.5	27.1
Paper-based Products	28	4.3	1.6	17.8	0.7	1.6	26.0
Natural Sands	11	22.5	0.3	0.6	1.4	0.1	25.0
Waste & Scrap	41	4.0	3.3	8.2	1.8	7.6	24.9
Misc. Manufactured	40	1.9	1.8	16.3	0.5	3.1	23.6
Textiles & Leathers	30	1.4	2.6	11.5	1.7	6.1	23.2
Machinery	34	2.8	6.0	4.4	5.2	2.2	20.6
Agricultural Products	03	5.9	4.8	6.3	2.7	0.7	20.5
Wood Products	26	3.9	4.2	4.1	4.2	3.8	20.2
Transport Equipment	37	13.7	1.3	1.1	1.6	0.6	18.4
Precision Instruments	38	5.5	1.8	5.9	1.0	3.9	18.0
Mineral Products	31	3.5	5.0	3.6	1.7	1.9	15.7
Crude Oil	16	7.0	0.0	2.8	0.0	5.3	15.1
Non-metallic Minerals	13	4.8	2.9	3.3	1.6	0.9	13.4
Base Metals	32	2.0	2.4	3.1	2.4	2.1	12.0
Newsprints & Papers	27	3.1	1.2	3.7	1.3	1.3	10.5
Fertilizers	22	6.0	2.4	0.4	0.6	0.7	10.1
Cereal Grains	02	5.0	2.5	0.6	1.1	0.5	9.8
Motorized Vehicles	36	0.6	1.3	5.5	1.0	0.3	8.8
Coal	15	0.3	1.8	1.4	0.4	4.5	8.3
Gravel & Crushed Stone	12	0.5	2.0	0.1	5.2	0.1	7.8
Logs	25	1.8	0.8	2.0	0.6	0.2	5.3
Metallic Ores	14	1.6	0.9	1.6	0.2	0.5	4.7
Live Animals & Fish	01	0.0	0.0	2.6	0.0	0.0	2.6
Tobacco Products	09	0.0	0.0	0.3	0.0	0.3	0.6

Table 5: Top Outlier MSAs Ranked in Order of Commodity Value Proportion Moved

Commodity	Top MSA by Value	Weight %	Value %
Alcoholic Beverages	Chicago, IL	61.5	56.3
Raw Meats	Los Angeles, CA	40.6	53.4
Gasoline	Houston, TX	46.8	47.3
Plastics & Rubbers	Houston, TX	40.0	40.0
Basic Chemicals	Houston, TX	36.3	38.1
Printed Products	Los Angeles, CA	35.8	38.0
Chemical Products	Houston, TX	33.7	36.7
Fuel Oils	Houston, TX	29.8	35.8
Furniture	Los Angeles, CA	33.4	35.4
Electronics	Los Angeles, CA	28.8	35.1
Products of Animals	Los Angeles, CA	25.4	35.0
Base Metal Articles	Houston, TX	39.1	33.5
Milled Grains	Newark, NJ	27.6	29.8
Petroleum Products	Houston, TX	21.3	28.3
Pharmaceuticals	Newark, NY	28.3	28.1
Prepared Foods	Los Angeles, CA	16.7	27.1
Misc. Manufactured	Los Angeles, CA	19.9	23.6
Textiles & Leathers	Los Angeles, CA	15.3	23.2
Base Metals	Chicago, IL	9.7	12.0

Each histogram points to the outlier proportions. For example, the proportion of alcoholic beverages moved by the Chicago MSA is more than four times that of the other items it moved. Similarly, the proportion of raw meats moved by the Los Angeles MSA was more than four times that of the other items it moved by rail.

This analysis suggests that any loss of railroad capacity can severely disrupt the supply chain for raw meats, alcoholic beverages, gasoline, and paper-based products such as toilet paper because of their extreme proportion by weight (more than 40%) moved by the accident outlier (high-risk) MSAs. As experienced during the COVID-19 pandemic, shortages due to supply chain disruptions can lead to price gouging and hoarding as consumers rush to acquire scarce items like toilet paper, tissues, and raw meats. Conversely, commodity categories with the least risk of shortages from railroad capacity disruptions at the accident outlier MSAs include tobacco products, live animals and fish, metallic ores, gravel and crushed stones, and logs.

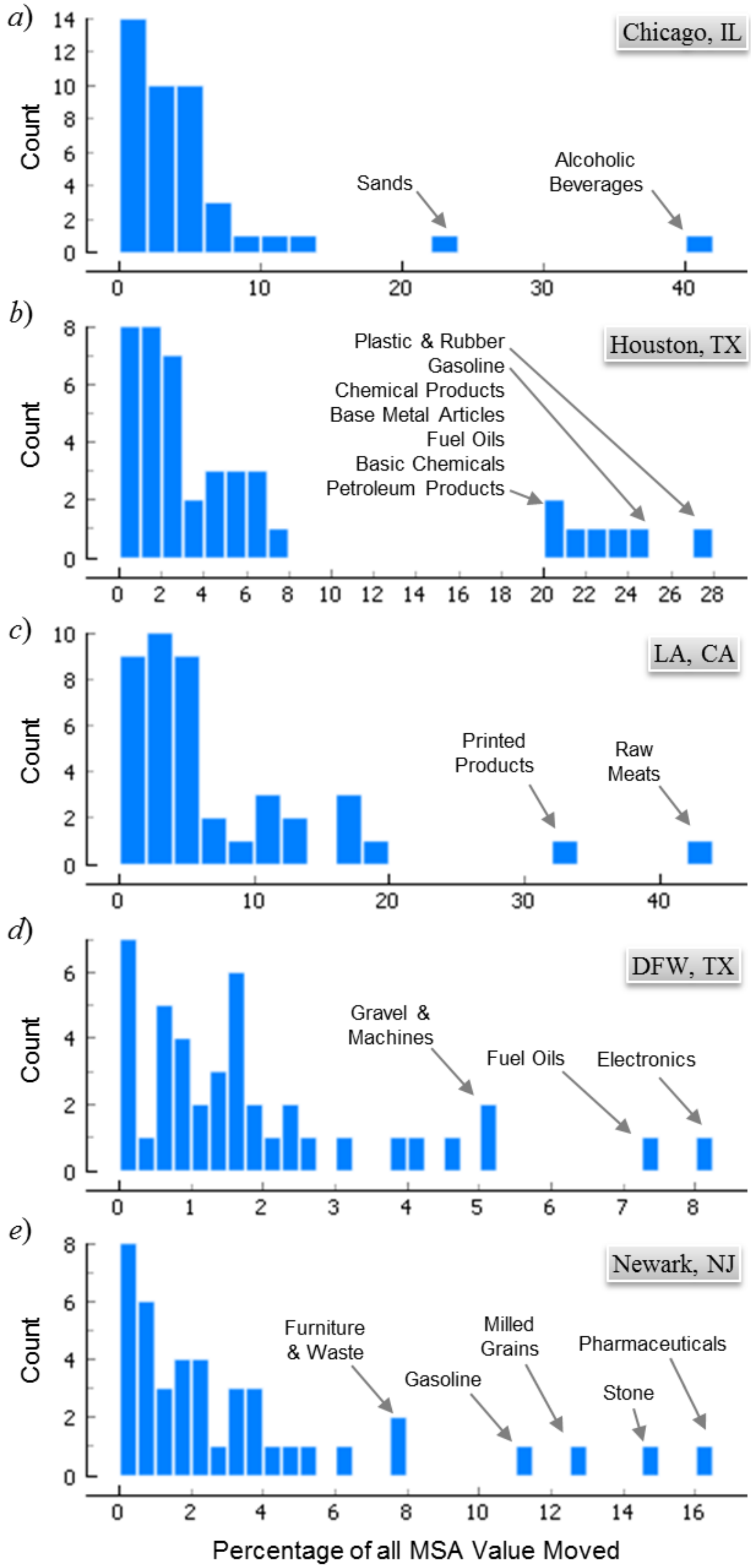

Figure 4: Value proportion distribution of commodity categories moved by the outlier MSAs.

Those items account for less than 5% of the commodity proportion by weight moved by the outlier (high-risk) MSAs. Therefore, supply chain managers can focus their efforts and resources on locations and commodities that are at the highest risk of flow disruption.

4 Discussions

The outlier MSAs contain the largest U.S. cities, which is consistent with the intuition that they will generate the highest demand to move products. Previous work found that there is a high correlation between train traffic and accident rates (Schafer and Barkan 2008). Therefore, a higher demand for moving products by rail increases the risk of accidents. Related work found that financial losses from railroad accidents tend to peak in the summer months, particularly due to the increase in shipping harvested gains (Dhingra, et al. 2022).

Implications from the findings of this study extend beyond the realm of supply chain management because of the profound impacts that railroad disruptions can have on society. The COVID-19 pandemic highlighted the fragile nature of global supply chains. An interruption in the supply of top commodities like alcoholic beverages and raw meats could lead to hoarding, price gouging, and widespread disruption in the restaurant and hospitality industries. Such actions could exacerbate social inequalities because lower-income households may not be able to afford inflated prices due to shortages. Small to medium enterprises would struggle to adapt in such situations, which could potentially lead to business closures and job losses. Interrupted supply chains can strain local economies and create a ripple effect that could extend beyond the disrupted areas. Those societal implications underscore the need for risk mitigation strategies.

The five MSAs identified as the most vulnerable provide clear targets for managing the risks of railroad disruptions. Initially focusing on railroad operations in those MSAs can help to prevent severe disruptions because they move more than 40% of the monetary value in certain commodities. Potential risk mitigation solutions could include diversifying transportation modes for the highest-risk commodities. For example, agencies can plan to use emerging modes such as electrified and autonomous trucks and electrified freight drones to help with transport diversification while reducing both cost and environmental harm.

Another potential solution would be to focus investments on improving infrastructure resilience and creating alternative rail routes to relieve high traffic nodes or bottleneck routes on the rail network. Other measures to improve resilience could include hardening the rail infrastructure to withstand extreme weather events and enhancing cybersecurity measures to protect connected railroads (using PTC) from cyberattacks.

Public policy can also play a crucial role in promoting risk management. For example, policymakers could incentivize the diversification of and hardening of rail transport by first targeting operations in the most vulnerable MSAs identified. Regulations to prevent price gouging in the event of supply chain disruptions would help to protect consumers and businesses.

Managers globally can use the data mining workflow as a tool to identify high-risk areas in their own supply chains and implement targeted risk mitigation strategies. In summary, proactive policymaking to diversify and harden freight transport could contribute to a more stable and resilient global supply chain that can withstand future disruptions to safeguard society and the economy.

5 Conclusions

The loss of human resources due to the COVID-19 pandemic sensitized the public to how any loss of transport capacity at freight bottlenecks like ports and inland terminals can cause widespread supply chain disruptions. Railroads are especially vulnerable to movement disruptions because the rail network has few alternative routes to accommodate detours in the event of disruptions. For instance, railroads are consistently involved in thousands of accidents annually that can disrupt the flow of certain commodities much more than others. Additional risk factors are threats such as extreme weather events, cyberattacks, and cargo theft.

This research developed a data mining workflow to rank the risk of commodity flow disruptions in the continental United States (CONUS). A key finding is that railroad accidents are at least five times more likely to occur in five MSAs of the CONUS than in other areas. Those five MSAs moved more than 40% of the monetary value in alcoholic beverages, raw meats, gasoline, plastic-based products, and rubber-based products. Alcoholic beverages are the top rail-transported commodity in the CONUS. The Chicago MSA accounted for more than 61% and 56% of the alcoholic beverages moved on rail by weight and monetary value, respectively. The Los Angeles and Houston MSAs are two accident outlier areas that handled 15 of the 42 commodity categories moved by rail. The above findings suggest that supply chain managers should focus risk mitigation and resiliency strategies in those five MSAs to minimize potential disruption that can lead to severe shortages of those commodities in the nation. For instance, a severe shortage of the two top commodities (alcoholic beverages and raw meats) can lead to hording, price gouging, and widespread disruptions in the restaurant and hospitality industry. Supply chain managers can use the data mining workflow developed in this research to rank commodity flow disruption risks in any region of the world, and to identify locations where to focus risk mitigation strategies.

Future work will apply a modification of the methodological workflow to profile the transport risks of commodities moved by airways and by trucks to determine if emerging freight drones could help to mitigate the risk of capacity losses from those modes.

6 References

Agresti, Alan. 2018. *Statistical Methods for the Social Sciences.* 5th. Boston, Massachusetts: Pearson.

Basallo-Triana, Mario José, Juan José Bravo-Bastidas, and Carlos Julio Vidal-Holguín. 2022. "A rail-road transshipment yard picture." *Transportation Research Part E: Logistics and Transportation Review* 159: 102629. doi:10.1016/j.tre.2022.102629.

Bešinović, Nikola. 2020. "Resilience in railway transport systems: a literature review and research agenda." *Transport Reviews* 40 (4). doi:10.1080/01441647.2020.1728419.

Bhardwaj, Bhavana, Raj Bridgelall, Pan Lu, and Neeraj Dhingra. 2021. "Signal Feature Extraction and Combination to Enhance the Detection and Localization of Railroad Track Irregularities." *IEEE Sensors Journal* 21 (5). doi:10.1109/JSEN.2020.3041652.

Bridgelall, Raj, and Denver D Tolliver. 2022. "Budgeting for the adoption of sensors on connected trains." *Transportation Planning and Technology* 45 (1): 1-17. doi:10.1080/03081060.2021.2017205.

Bridgelall, Raj, and Denver D. Tolliver. 2021. "Railroad accident analysis using extreme gradient boosting." *Accident Analysis and Prevention* 156. doi:10.1016/j.aap.2021.106126.

BTS and USCB. 2020. *2017 Commodity Flow Survey Methodology.* Washington, DC: U.S. Department of Transportation, Bureau of Transportation Statistics (BTS), and U.S. Department of Commerce, U.S. Census Bureau, 28. https://www2.census.gov/programs-surveys/cfs/technical-documentation/methodology/2017cfsmethodology.pdf.

BTS and USCB. 2015. *2017 Commodity Flow Survey Standard Classification of Transported Goods (SCTG).* SCTG Commodity Codes, Washington, DC: Bureau of Transportation Statistics and U.S. Census Bureau. https://www2.census.gov/programs-surveys/cfs/technical-documentation/code-list/CFS-1200_17.pdf.

Darayi, Mohamad, Kash Barker, and Charles D. Nicholson. 2019. "A multi-industry economic impact perspective on adaptive capacity planning in a freight transportation network." *International Journal of Production Economics* 208. doi:10.1016/j.ijpe.2018.12.008.

Dhingra, Neeraj, Raj Bridgelall, Pan Lu, Joseph Szmerekovsky, and Bhavana Bhardwaj. 2022. "Ranking Risk Factors in Financial Losses from Railroad Incidents: A Machine Learning Approach." *Transportation Research Record: Journal of the Transportation Research Board* 2677 (2). doi:10.1177/03611981221133085.

Duggan, Sarah, and Mark McMurtrey. 2021. "Distribution in the Food Industry: Impact of the 2020 Truck Driver Shortage." *Journal of Marketing Development and Competitiveness* 15 (2). doi:10.33423/jmdc.v15i2.4333.

FHWA. 2021. *Freight Analysis Framework Verion 5 (FAF5).* November 22. Accessed February 13, 2022. https://faf.ornl.gov/faf5/Default.aspx.

FRA. 2021. *Federal Railroad Administration.* July 28. https://safetydata.fra.dot.gov/OfficeofSafety/default.aspx.

—. 2021. *Monetary Threshold Notice.* Federal Railroad Administration (FRA). 15 December. Accessed February 16, 2022. https://railroads.dot.gov/forms-guides-publications/guides/monetary-threshold-notice.

Ghofrani, Faeze, Qing He, Rob M.P. Goverde, and Xiang Liu. 2018. "Recent applications of big data analytics in railway transportation systems: A survey." *Transportation Research Part C: Emerging Technologies* 90. doi:10.1016/j.trc.2018.03.010.

Huang, Ping, Chao Wen, Liping Fu, Javad Lessan, Chaozhe Jiang, Qiyuan Peng, and Xinyue Xu. 2020. "Modeling train operation as sequences: A study of delay prediction with operation and weather data." *Transportation Research Part E: Logistics and Transportation Review* 141: 102022. doi:10.1016/j.tre.2020.102022.

Jabbarzadeh, Armin, Nader Azad, and Manish Verma. 2020. "An optimization approach to planning rail hazmat shipments in the presence of random disruptions." *Omega (United Kingdom)* 96. doi:10.1016/j.omega.2019.06.004.

Ke, Ginger Y., and Manish Verma. 2021. "A framework to managing disruption risk in rail-truck intermodal transportation networks." *Transportation Research Part E: Logistics and Transportation Review* 153. doi:10.1016/j.tre.2021.102340.

Kelle, Peter, Jinglu Song, Mingzhou Jin, Helmut Schneider, and Christopher Claypool. 2019. "Evaluation of operational and environmental sustainability tradeoffs in multimodal freight transportation planning." *International Journal of Production Economics* 209. doi:10.1016/j.ijpe.2018.08.011.

Kolli, Satish, Joshua Lilly, and Duminda Wijesekera. 2018. "Positive train control security: An intrusion-detection system to provide cyber-situational awareness." *IEEE Vehicular Technology Magazine* 13 (3). doi:10.1109/MVT.2018.2848398.

Kour, Ravdeep, Ramin Karim, and Adithya Thaduri. 2020. "Cybersecurity for railways – A maturity model." *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* 234 (10). doi:10.1177/0954409719881849.

Lorenc, Augustyn, and Małgorzata Kuźnar. 2021. "The most common type of disruption in the supply chain - Evaluation based on the method using artificial neural networks." *International Journal of Shipping and Transport Logistics* 13 (1-2). doi:10.1016/j.trc.2017.09.009.

Lorenc, Augustyn, Małgorzata Kuźnar, Tone Lerher, and Maciej Szkoda. 2020. "Predicting the probability of cargo theft for individual cases in railway transport." *Tehnicki Vjesnik* 27 (3). doi:10.17559/TV-20190320194915.

Perrow, Charles. 1999. "Organizing to reduce the vulnerabilities of complexity." *Journal of Contingencies and Crisis Management* 7 (3). doi:10.1111/1468-5973.00108.

Procházka, J., Š. Hošková-Mayerová, and D. Procházková. 2020. "The risks connected with accidents on highways and railways." *Quality & Quantity* 54 (5): 1537-1548. doi:10.1007/s11135-019-00899-1.

Schafer, Darwin H., and Christopher PL Barkan. 2008. "Relationship Between Train Length and Accident Causes and Rates." *Transportation Research Record: Journal of the Transportation Research Board* 2043 (1): 73-82. doi:10.3141/2043.

Scheibe, Kevin P., and Jennifer Blackhurst. 2018. "Supply chain disruption propagation: a systemic risk and normal accident theory perspective." *International Journal of Production Research* 56 (1-2). doi:10.1080/00207543.2017.1355123.

Schofer, Joseph L, Hani S Mahmassani, and Max T M Ng. 2022. "Resilience of US Rail Intermodal Freight during the Covid-19 Pandemic." *Research in Transportation Business & Management* 100791. doi:10.1016/j.rtbm.2022.100791.

Sharma, Sunil Kumar, Sanghmitra Poddar, G.K. Dwivedy, Subhash Chandra Panja, and S.N. Patra. 2021. "Risk reduction and resilience buildup in railroad transport." In *Disaster Resilience and Sustainability: Adaptation for Sustainable Development*, edited by Indrajit Pal, Riyanti Djalante, Rajib Shaw and Sangam Shrestha. Elsevier Inc. doi:10.1016/B978-0-323-85195-4.00033-0.

Thaduri, Adithya, Mustafa Aljumaili, Ravdeep Kour, and Ramin Karim. 2019. "Cybersecurity for eMaintenance in railway infrastructure: risks and consequences." *International Journal of Systems Assurance Engineering and Management* 10 (2). doi:10.1007/s13198-019-00778-w.

Uddin, Majbah, and Nathan Huynh. 2019. "Reliable Routing of Road-Rail Intermodal Freight under Uncertainty." *Networks and Spatial Economics* 19 (3). doi:10.1007/s11067-018-9438-6.

USCB. 2021. *Commodity Flow Survey Geographies 2017.* October 8. Accessed February 13, 2022. https://www.census.gov/programs-surveys/cfs/technical-documentation/geographies.html.

Wendler-Bosco, Vera, and Charles Nicholson. 2020. "Port disruption impact on the maritime supply chain: a literature review." *Sustainable and Resilient Infrastructure* 5 (6). doi:10.1080/23789689.2019.1600961.

Woodburn, Allan. 2019. "Rail network resilience and operational responsiveness during unplanned disruption: A rail freight case study." *Journal of Transport Geography* 77. doi:10.1016/j.jtrangeo.2019.04.006.

Young, Jessica. 2022. "Early estimates: US ecommerce grows 14.2% in 2021." *digitalcommerce360.com*, February 10. https://www.digitalcommerce360.com/article/us-ecommerce-sales/.

Zhang, Zhipeng, Xiang Liu, and Keith Holt. 2018. "Positive Train Control (PTC) for railway safety in the United States: Policy developments and critical issues." *Utilities Policy* 51 (2018): 33-40.

Zhong, Ray Y., Xun Xu, and Olga Battaïa. 2020. "Special issue on sustainability with innovation for manufacturing and supply chain management." *International Journal of Production Research* 58 (24): 7311-7313. doi:10.1080/00207543.2020.1813466.

Figure 1 Caption: MSAs of the CONUS and the major railroad routes that serve them.

Figure 1 Alt Text: A map of the continental United States with lines that indicate the location and extent of railroad tracks, overlayed with colored polygons that indicate the locations and relative sizes of metropolitan areas that handle the freight analyzed in this study.

Figure 2 Caption: Input datasets and analytical workflow.

Figure 2 Alt Text: A three-section flow diagram, with each section containing labeled boxes and directional arrows between them to indicate the data-mining workflow and analysis.

Figure 3 Caption: Distribution of the number of railroad accidents accumulated from 2009 to 2019.

Figure 3 Alt Text: A combination bar and line chart with the horizontal axis indicating data bins for the quantity of accidents and the height of each bar (vertical axis) indicating the number of metropolitan areas where that volume of accidents occurred. The line overlays the bars to show the best fit lognormal distribution that excludes the outliers. Five labeled boxes contain the name of the outlier metropolitan areas and point to the corresponding outlier bars.

Figure 4 Caption: Value proportion distribution of commodity categories moved by the outlier MSAs.

Figure 4 Alt Text: Five separate bar charts (each representing one of the five outlier metropolitan areas) vertically stacked to compare the distribution of all commodity categories moved by value proportion. Labels and arrows point to outlier bars that indicate the type of commodity mostly moved by each outlier metropolitan area.