Project title

Proposal

marvelous echidna

```
library(tidyverse)
library(skimr)
library(readxl)
```

Data 1

Introduction and data

- Identify the source of the data.

https://essd.copernicus.org/articles/13/5213/2021/

https://publications.jrc.ec.europa.eu/repository/handle/JRC125851

https://zenodo.org/records/6483002

- State when and how it was originally collected (by the original data curator, not necessarily how you found the data).

Their paper was written in 2021 and published in 2022. The dataset on anthropogenic greenhouse gas emissions covering the years 1970-2019 was originally collected by the "Emissions Database for Global Atmospheric Research" (EDGAR), and used by this team to identify the greenhouse gas emissions. The EDGAR database is a well-known and widely used resource for greenhouse gas emissions data, providing comprehensive information on emissions from various sources, including fossil fuel combustion and industry, methane emissions, nitrous oxide emissions, fluorinated gases, and net CO2 emissions from land use, land-use change, and forestry (LULUCF). The data from EDGAR are collected through a combination of methods, including direct measurements, statistical modeling, and estimation based on activity data and emission factors. The dataset is continually updated and refined by the EDGAR project team to provide accurate and reliable information on greenhouse gas emissions globally. The dataset from the source that used EDGAR's data also include information from other sources, such as

bookkeeping models for net CO2 emissions from LULUCF, to complement and enhance the coverage and accuracy of the emissions data.

- Write a brief description of the observations.

The research data highlights discrepancies in global greenhouse gas emissions inventories, emphasizes the need for improved reporting methods, and underscores the importance of high-quality data for tracking climate change progress. The dataset, sourced from EDGAR and bookkeeping models, provides comprehensive information on global GHG emissions by sector from 1970 to 2019, revealing trends in emissions reductions and increases across regions and economic sectors. Some of the trends include: Sustained emissions growth: There is a pattern of sustained emissions growth over the five decades analyzed, with high confidence that global anthropogenic GHG emissions have increased every decade; and, Dominance of CO2 emissions: CO2 has accounted for almost 75% of the emissions growth since 1970 in terms of CO2 equivalent, making it the dominant driver of emissions growth.

Research question

- A well formulated research question. (You may include more than one research question if you want to receive feedback on different ideas for your project. However, one per data set is required.)

What is the relationship between the energy sector's greenhouse gas emissions and economic development in different regions from 1970 to 2019?

- A description of the research topic along with a concise statement of your hypotheses on this topic.

Understanding the relationship between greenhouse gas emissions from the energy sector and economic development is crucial for informing sustainable development policies. By examining how economic growth influences emissions and vice versa, policymakers can design strategies to mitigate climate change while promoting economic prosperity.

Hypothesis: Higher levels of economic development will be associated with increased greenhouse gas emissions from the energy sector due to greater energy demand and industrial activities.

- Identify the types of variables in your research question. Categorical? Quantitative?

Dependent variable: Greenhouse gas emissions from the energy sector (quantitative). Independent variable: Economic development indicators such as GDP per capita or industrial output (quantitative). Moderator variables: Region (categorical), time period (1970-2019, quantitative). Control variables: Energy efficiency measures, renewable energy adoption, policy interventions, population growth, and technological advancements.

Glimpse of data

```
# add code here
ghg_data_gwp100 <- read_excel("data/essd_ghg_data_gwp100.xlsx",
    sheet = "data")
lulucf_data <- read_excel("data/essd_lulucf_data.xlsx",
    sheet = "data")
skimr::skim(ghg_data_gwp100)
```

Table 1: Data summary

Name	ghg_data_gwp100
Number of rows	196639
Number of columns	16
_____	
Column type frequency:	
character	10
numeric	6
_____	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ISO	0	1	3	3	0	230	0
country	0	1	4	44	0	230	0
region_ar6_6	0	1	6	36	0	8	0
region_ar6_6_short	0	1	3	3	0	8	0
region_ar6_10	0	1	6	36	0	12	0
region_ar6_10_short	0	1	3	16	0	12	0
region_ar6_22	0	1	9	36	0	21	0
region_ar6_dev	0	1	3	14	0	5	0
sector_title	0	1	5	14	0	5	0
subsector_title	0	1	4	38	0	26	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
year	0	1	1995.17	14.65	1970.0	1983.00	1995.00	2008.00	2020	
CO2	0	1	6680859.22	24344331.01	0.0	0.00	22569.20	750614.70	1794320934	
CH4	0	1	2143170.96	16094553.23	-749415.8	3.04	2267.27	98913.77	661000323	
N2O	0	1	538887.27	4364087.28	0.0	4.66	2141.34	50244.83	156637863	
Fgas	0	1	148110.67	2857963.87	0.0	0.00	0.00	0.00	217895030	
GHG	0	1	9511028.63	37851192.06	-483078.7	28800.82	350835.02	2784196.47	1794320934	

```
skimr::skim(lulucf_data)
```

Table 4: Data summary

Name	lulucf_data
Number of rows	1700
Number of columns	9
_____	
Column type frequency:	
character	4
numeric	5
_____	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
region_ar6_6	0	1	6	36	0	6	0
region_ar6_6_short	0	1	3	3	0	6	0
region_ar6_10	0	1	6	36	0	10	0
region_ar6_10_short	0	1	6	16	0	10	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
year	0	1	1934.5	49.09	1850	1892	1934.5	1977	2019	
blue	0	1	591751860	910944043.51	-267956839	1608413.75	38468803.5	516749120	2024670224	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
houghton	0	1	32062249	228227293.26	-674291031	1724543	8561563	39741201	4000301670	
oscar	0	1	43335442	158770793.03	-321211980	829677	7310058	28589423	4387408072	
mean	0	1	44857626	232775171.35	-257943381	953278	8327193	46408347	3466350873	

Data 2

Introduction and data

- Identify the source of the data.

The World bank. https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators

- State when and how it was originally collected (by the original data curator, not necessarily how you found the data).

The data is annually collected and compiled by the World Bank from officially-recognized international sources, offering national, regional, and global estimates that represent the most current and accurate global development data available.

- Write a brief description of the observations.

Observations in this dataset include annual data for over 200 countries across various indicators such as health, education, economic performance, infrastructure, and environmental metrics. Each observation provides insight into the status and progress of a country on these indicators for a given year. ## Research question

- A well formulated research question. (You may include more than one research question if you want to receive feedback on different ideas for your project. However, one per data set is required.)

How have health expenditures and life expectancy changed over the past two decades, and what is their relationship with economic performance across countries?

What is the impact of access to improved water sources and sanitation facilities on child mortality rates in low and lower-middle-income countries over the last two decades?

- A description of the research topic along with a concise statement of your hypotheses on this topic.

This research aims to investigate the relationship between public health infrastructure—specifically, access to improved water sources and sanitation facilities—and child mortality rates. The underlying hypothesis is that increased access to improved water sources and sanitation facilities is associated with lower child mortality rates. This relationship is expected to be more pronounced in low and lower-middle-income countries, where improvements in basic infrastructure can significantly affect health outcomes.

- Identify the types of variables in your research question. Categorical? Quantitative?

Health Expenditure (% of GDP): Total expenditure on health as a percentage of GDP - Quantitative Life Expectancy at Birth (Years): Average number of years a newborn is expected to live - Quantitative GDP Growth (Annual %): Annual percentage increase in a country's GDP - Quantitative Country Income Group: Classification based on GDP per capita (High income, Upper middle income, Lower middle income, Low income) - Categorical Access to Improved Water Sources (% of Population): Percentage with access to safe drinking water - Quantitative Access to Improved Sanitation Facilities (% of Population): Percentage with access to improved sanitation - Quantitative Under-5 Mortality Rate (Deaths per 1,000 Live Births): Deaths of children under five years old per 1,000 live births - Quantitative Mortality Rate Attributed to Unsafe Water, Unsafe Sanitation, and Lack of Hygiene (Per 100,000 Population): Mortality rate due to lack of access to safe water, sanitation, and hygiene - Quantitative

Glimpse of data

```
file_path <- "data/WDIEXCEL.xlsx"
wdi_data <- read_excel(file_path)
skim(wdi_data)
```

Table 7: Data summary

Name	wdi_data
Number of rows	395276
Number of columns	67
_____	
Column type frequency:	
character	4
numeric	63
_____	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Country Name	0	1	4	52	0	266	0
Country Code	0	1	3	3	0	266	0
Indicator Name	0	1	10	141	0	1486	0
Indicator Code	0	1	6	25	0	1486	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
1960	356394	0.10	4.353165e+13	1.981070e+13	- 2.594795e+14	4.38	34.17	50195.25	1.629367e+15	
1961	351129	0.11	4.103759e+13	2.015732e+13	- 3.243188e+14	4.44	35.22	42904.50	1.798673e+15	
1962	349947	0.11	4.225246e+13	2.118711e+13	- 3.675279e+14	4.50	33.94	38998.00	1.940935e+15	
1963	349067	0.12	4.324092e+13	2.222237e+13	- 4.061755e+14	4.57	33.65	38431.00	2.077617e+15	
1964	348524	0.12	4.637837e+13	2.397489e+13	- 4.462457e+14	4.65	34.40	44349.50	2.253700e+15	
1965	346318	0.12	5.074775e+13	2.715842e+13	- 5.422808e+14	4.49	35.92	72000.00	2.637633e+15	
1966	346316	0.12	5.516202e+13	3.006712e+13	- 6.431337e+14	4.74	37.53	85431.75	2.941025e+15	
1967	345648	0.13	6.031129e+13	3.335017e+13	- 7.339951e+14	4.72	38.62	110000.00	3.271972e+15	
1968	344995	0.13	6.792608e+13	3.783570e+13	- 8.208008e+14	4.85	39.65	130000.00	3.743273e+15	
1969	344232	0.13	7.518070e+13	4.296538e+13	- 1.029098e+15	4.91	40.43	160000.00	4.324104e+15	
1970	326876	0.17	6.490753e+13	4.117363e+13	- 1.193370e+15	4.73	41.40	422286.27	4.796625e+15	
1971	319944	0.19	6.762925e+13	4.509756e+13	- 1.031207e+15	4.81	39.00	150953.25	5.453839e+15	
1972	318159	0.20	7.579660e+13	5.144161e+13	- 1.176888e+15	5.00	39.86	201000.00	6.244592e+15	
1973	317475	0.20	8.826931e+13	5.884442e+13	- 1.842127e+14	5.09	40.53	245720.00	7.011691e+15	
1974	316062	0.20	1.177419e+14	8.027756e+13	- 1.316864e+14	5.18	41.73	315070.50	1.201491e+16	
1975	312342	0.21	1.195363e+14	8.073605e+13	- 1.321367e+14	4.97	40.95	380389.25	1.136509e+16	

skim_variable	missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
1976	310226	0.22	1.330209e+13	9.194432e+13	- 1.491390e+14	5.21	41.84	504497.24	1.320145e+16	
1977	307819	0.22	1.286832e+13	8.829624e+13	- 1.654078e+14	5.28	42.75	727750.00	1.249774e+16	
1978	307060	0.22	1.116652e+13	7.321734e+13	- 1.650252e+14	5.34	43.29	829999.98	1.002291e+16	
1979	306088	0.23	1.059618e+13	6.893019e+13	- 8.260263e+13	5.44	43.52	911057.25	1.066946e+16	
1980	301550	0.24	8.364066e+12	5.120779e+13	- 1.603200e+13	5.66	45.05	971402.75	7.221394e+15	
1981	299560	0.24	8.045148e+12	4.808626e+13	- 1.597800e+13	5.39	43.50	899999.98	6.796725e+15	
1982	299203	0.24	8.871038e+12	5.117055e+13	- 1.679300e+13	5.37	44.24	1071412.00	7.256677e+15	
1983	298217	0.25	1.109893e+13	6.728556e+13	- 1.809700e+13	5.48	44.70	1087361.75	9.976014e+15	
1984	297576	0.25	1.131585e+13	7.029362e+13	- 1.631300e+13	5.60	44.77	1118229.80	9.980000e+15	
1985	296134	0.25	1.228772e+13	8.280643e+13	- 1.485300e+13	5.55	45.25	1136858.25	1.251000e+16	
1986	295361	0.25	1.145162e+13	7.622552e+13	- 1.688300e+13	5.56	45.07	1159063.80	1.014000e+16	
1987	295145	0.25	1.155591e+13	7.053825e+13	- 1.987900e+13	5.56	45.05	1280515.25	1.014000e+16	
1988	295050	0.25	8.974620e+12	4.684445e+13	- 9.764782e+13	5.69	45.65	1481500.63	6.204935e+15	
1989	293295	0.26	9.731472e+12	5.037434e+13	- 2.583300e+13	5.74	46.76	1568683.00	6.507206e+15	
1990	269790	0.32	1.011441e+13	5.845122e+13	- 3.052100e+13	5.07	46.99	680000.01	7.910000e+15	
1991	259464	0.34	9.863369e+12	5.766697e+13	- 2.193638e+13	5.65	45.73	195967.86	7.673360e+15	
1992	256136	0.35	1.004365e+13	5.867988e+13	- 2.789349e+13	5.84	46.22	205936.75	7.597248e+15	
1993	253617	0.36	1.087615e+13	6.427326e+13	- 1.983469e+13	5.90	46.62	263377.59	9.208374e+15	
1994	251365	0.36	1.166083e+13	6.891360e+13	- 3.859458e+13	6.00	47.26	330364.50	1.102089e+16	
1995	243418	0.38	1.176225e+13	6.876676e+13	- 5.230384e+13	6.23	48.41	527668.75	1.195209e+16	

skim_variable	missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
1996	236240	0.40	1.108989e+13	5.218677e+13	- 2.983901e+13	5.81	45.06	230372.56	8.825688e+15	
1997	242867	0.39	1.178853e+13	5.322649e+13	- 2.552600e+13	6.21	48.44	655244.06	8.305691e+15	
1998	232832	0.41	1.026355e+13	5.196843e+13	- 3.291633e+15	5.64	44.87	201047.25	7.739116e+15	
1999	234118	0.41	1.056206e+13	5.361292e+13	- 4.958967e+15	6.24	48.01	284117.06	8.838588e+15	
2000	199330	0.50	9.245861e+13	5.194178e+13	- 7.109153e+15	5.63	40.44	9702.75	9.191774e+15	
2001	213223	0.46	1.040853e+13	6.155234e+13	- 8.938409e+15	6.25	44.56	38784.77	9.489688e+15	
2002	203258	0.49	1.200453e+13	6.147364e+13	- 2.794935e+15	5.94	41.85	18762.04	1.107315e+16	
2003	200934	0.49	1.170523e+13	7.561710e+13	- 9.824821e+15	6.00	42.14	18375.25	1.194001e+16	
2004	195654	0.51	1.327572e+13	7.387025e+13	- 5.361126e+15	6.11	41.66	12252.41	1.288384e+16	
2005	183680	0.54	1.354375e+13	7.122421e+13	- 7.157321e+15	6.00	39.25	5369.41	1.448172e+16	
2006	183309	0.54	1.503746e+13	8.246474e+13	- 6.200057e+15	6.00	39.72	6924.00	1.529797e+16	
2007	181305	0.54	1.778992e+13	8.089911e+13	- 1.691820e+14	5.92	40.00	7790.94	1.706973e+16	
2008	180495	0.54	1.956165e+13	9.319891e+13	- 4.636015e+14	5.98	40.67	9323.00	1.679243e+16	
2009	179217	0.55	2.018296e+13	9.428623e+13	- 1.716620e+14	5.90	40.33	8589.12	1.601307e+16	
2010	168121	0.57	2.257578e+13	1.053176e+14	- 1.827709e+14	5.66	38.06	4213.00	1.769461e+16	
2011	172345	0.56	2.568456e+13	1.128642e+14	- 2.168927e+14	6.02	40.81	6678.18	1.878185e+16	
2012	170613	0.57	2.679878e+13	1.168195e+14	- 2.431930e+14	5.75	39.79	6519.37	1.518257e+16	
2013	173335	0.56	2.970408e+13	1.279122e+14	- 2.853262e+14	6.00	40.88	9121.00	1.462878e+16	
2014	166314	0.58	3.121972e+13	1.370795e+14	- 3.543931e+14	5.79	40.28	4924.33	1.464430e+16	
2015	165938	0.58	3.196345e+13	1.399065e+14	- 3.823398e+14	5.82	40.91	7247.50	1.325074e+16	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
2016	170251	0.57	3.570034e+15	2.239230e+14	-3.970130e+14	5.71	40.70	7032.96	1.442551e+16	
2017	173908	0.56	3.832511e+15	1.923354e+14	-4.306058e+14	6.00	42.95	9392.75	1.675691e+16	
2018	175268	0.56	4.303226e+15	1.927117e+14	-6.109890e+14	5.56	39.84	8649.34	2.173981e+16	
2019	180043	0.54	2.984216e+16	1.152313e+16	-1.155334e+15	5.98	41.36	11970.00	5.391345e+18	
2020	199710	0.49	4.225128e+16	1.841788e+17	-2.023638e+15	5.56	44.89	43918.50	8.144900e+19	
2021	228925	0.42	2.601680e+17	1.057423e+18	-8.595719e+14	6.93	52.72	534427.86	4.312820e+20	
2022	280236	0.29	2.682362e+17	9.092173e+18	-9.091425e+14	6.04	45.28	2281091.49	3.083842e+21	

Data 3

Introduction and data

- Identify the source of the data.
  - https://catalog.data.gov/dataset/global-landslide-catalog-export
- State when and how it was originally collected (by the original data curator, not necessarily how you found the data).
  - The data was collected by searching for news reports, scientific reports, eyewitness statements, aerial photography, as well as other media that reliably reported the details of the landslide event (source: https://doi.org/10.1007/s11069-009-9401-4).
- Write a brief description of the observations.
  - The database records each landslide event that has been reported since 2007 until March 2016. The database records observations such as the date, time, trigger, fatalities, geographic location, as well as a host of other useful information.

Research question

- A well formulated research question. (You may include more than one research question if you want to receive feedback on different ideas for your project. However, one per data set is required.)

  - Are there areas that are more prone to experiencing landslides? If so, what types of areas are more vulnerable to landslides? Are landslides more deadly in different areas around the world? Have landslides increased in frequency?

- A description of the research topic along with a concise statement of your hypotheses on this topic.
  - Research into the global presence of landslides and potential clusters.
  - Hypothesis: Landslides are not a random occurrence, and predominantly occur in certain areas around the globe. These areas may have shared properties such as annual rainfall, soil composition, and perhaps even latitude.

- Identify the types of variables in your research question. Categorical? Quantitative?
  - A mix of qualitative and quantitative. Variables like location, landslide trigger, and country are qualitative, but variables like fatalities and date/time are quantitative.

Glimpse of data

```
data <- read_csv('data/Global_Landslide_Catalog_Export.csv')
```

```
Rows: 11033 Columns: 31
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (22): source_name, source_link, event_date, event_title, event_descripti...
dbl  (8): event_id, fatality_count, injury_count, event_import_id, admin_div...
lgl  (1): event_time

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
glimpse(data)
```

```
Rows: 11,033
Columns: 31
$ source_name              <chr> "AGU", "Oregonian", "CBS News", "Reuters", "~
$ source_link              <chr> "https://blogs.agu.org/landslideblog/2008/10~
$ event_id                 <dbl> 684, 956, 973, 1067, 2603, 4203, 4290, 225, ~
$ event_date               <chr> "08/01/2008 12:00:00 AM", "01/02/2009 02:00:~
$ event_time               <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
$ event_title              <chr> "Sigou Village, Loufan County, Shanxi Provin~
```

```
$ event_description          <chr> "occurred early in morning, 11 villagers bur~
$ location_description       <chr> "Sigou Village, Loufan County, Shanxi Provin~
$ location_accuracy          <chr> "unknown", "5km", "10km", "unknown", "5km", ~
$ landslide_category         <chr> "landslide", "mudslide", "landslide", "lands~
$ landslide_trigger          <chr> "rain", "downpour", "downpour", "monsoon", "~
$ landslide_size             <chr> "large", "small", "large", "medium", "medium~
$ landslide_setting          <chr> "mine", "unknown", "unknown", "unknown", "un~
$ fatality_count             <dbl> 11, 0, 10, 1, 0, 0, 0, 3, NA, 2, 4, 0, 0, 0,~
$ injury_count               <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
$ storm_name                 <chr> NA, NA, NA, NA, "Supertyphoon Juan (Megi)", ~
$ photo_link                 <chr> NA, NA, NA, NA, NA, NA, NA, "http://www.cron~
$ notes                      <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
$ event_import_source        <chr> "glc", "glc", "glc", "glc", "glc", "glc", "g~
$ event_import_id            <dbl> 684, 956, 973, 1067, 2603, 4203, 4290, 225, ~
$ country_name               <chr> "China", "United States", "Peru", "Nepal", "~
$ country_code               <chr> "CN", "US", "PE", "NP", "PH", "PH", "US", "M~
$ admin_division_name        <chr> "Shaanxi", "Oregon", "Junín", "Mid Western",~
$ admin_division_population <dbl> 0, 36619, 14708, 20908, 798634, 2404, 2126, ~
$ gazeteer_closest_point     <chr> "Jingyang", "Lake Oswego", "San Ramón", "Dai~
$ gazeteer_distance          <dbl> 41.02145, 0.60342, 0.85548, 0.75395, 2.02204~
$ submitted_date             <chr> "04/01/2014 12:00:00 AM", "04/01/2014 12:00:~
$ created_date               <chr> "11/20/2017 03:17:00 PM", "11/20/2017 03:17:~
$ last_edited_date           <chr> "02/15/2018 03:51:00 PM", "02/15/2018 03:51:~
$ longitude                  <dbl> 107.4500, -122.6630, -75.3587, 81.7080, 123.~
$ latitude                   <dbl> 32.56250, 45.42000, -11.12950, 28.83780, 10.~
```

```
skim(data)
```

Table 10: Data summary

Name	data
Number of rows	11033
Number of columns	31
Column type frequency:	
character	22
logical	1
numeric	8
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
source_name	0	1.00	2	172	0	3918	0
source_link	846	0.92	23	386	0	8294	0
event_date	0	1.00	22	22	0	6550	0
event_title	0	1.00	3	150	0	10546	0
event_description	862	0.92	3	1003	0	9401	0
location_description	102	0.99	3	412	0	10432	0
location_accuracy	2	1.00	3	7	0	9	0
landslide_category	1	1.00	5	19	0	14	0
landslide_trigger	23	1.00	4	23	0	18	0
landslide_size	9	1.00	5	12	0	6	0
landslide_setting	69	0.99	4	16	0	14	0
storm_name	10456	0.05	1	41	0	217	0
photo_link	9537	0.14	28	292	0	1469	0
notes	10716	0.03	13	484	0	265	0
event_import_source	1563	0.86	3	80	0	3	0
country_name	1562	0.86	4	32	0	141	0
country_code	1564	0.86	2	2	0	139	0
admin_division_name	1637	0.85	3	36	0	887	0
gazeteer_closest_point	1563	0.86	2	45	0	4389	0
submitted_date	10	1.00	22	22	0	3786	0
created_date	1	1.00	22	22	0	419	0
last_edited_date	0	1.00	22	22	0	1	0

Variable type: logical

skim_variable	n_missing	complete_rate	mean	count
event_time	11033	0	NaN	:

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
event_id	0	1.00	5598.95	3249.23	1.00	2785.00	5563.00	8435.00	11221.00	
fatality_count	1385	0.87	3.22	59.89	0.00	0.00	0.00	1.00	5000.00	
injury_count	5674	0.49	0.75	8.46	0.00	0.00	0.00	0.00	374.00	
event_import_id	1562	0.86	4798.56	2789.13	-111.17	2386.50	4773.00	7189.50	9669.00	
admin_division_population	1562	0.86	157760.05	529734.50	0.00	1963.00	7365.00	34021.00	12691836.00	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
gazeteer_distance	1562	0.86	11.87	15.60	0.00	2.36	6.25	15.82	215.45	
longitude	0	1.00	2.52	100.91	-179.98	-107.87	19.69	93.95	179.99	
latitude	0	1.00	25.88	20.42	-46.77	13.92	30.53	40.87	72.63	