

# Energy Consumption Prediction

Climate change is a globally relevant, urgent, and multi-faceted issue heavily impacted by energy policy and infrastructure. Addressing climate change involves mitigation (i.e. mitigating greenhouse gas emissions) and adaptation (i.e. preparing for unavoidable consequences). Mitigation of GHG emissions requires changes to electricity systems, transportation, buildings, industry, and land use.

According to a report issued by the International Energy Agency (IEA), the lifecycle of buildings from construction to demolition were responsible for 37% of global energy-related CO2 emissions in 2020. Yet it is possible to drastically reduce the energy consumption of buildings by a combination of easy-to-implement fixes and state-of-the-art strategies. For example, retrofitted buildings can reduce heating and cooling energy requirements by 50-90 percent. Many of these energy efficiency measures also result in overall cost savings and yield other benefits, such as cleaner air for occupants. This potential can be achieved while maintaining the services that buildings provide.

## Goal

The goal of this competition is to predict the energy consumption using building characteristics and climate and weather variables.

## Data

The WiDS Datathon 2022 focuses on a prediction task involving roughly 100k observations of building energy usage records collected over 7 years and a number of states within the United States. The dataset consists of building characteristics (e.g. floor area, facility type etc), weather data for the location of the building (e.g. annual average temperature, annual total precipitation etc) as well as the energy usage for the building and the given year, measured as Site Energy Usage Intensity (Site EUI). Each row in the data corresponds to the a single building observed in a given year. Your task is to predict the Site EUI for each row, given the characteristics of the building and the weather data for the location of the building.

## Setup

### Load Data and Packages

```
library(gtExtras)
library(gtsummary)
library(tidyverse)
library(tidymodels)

data <- read_csv("data/buildings-data/train.csv")
```

### Clean Data

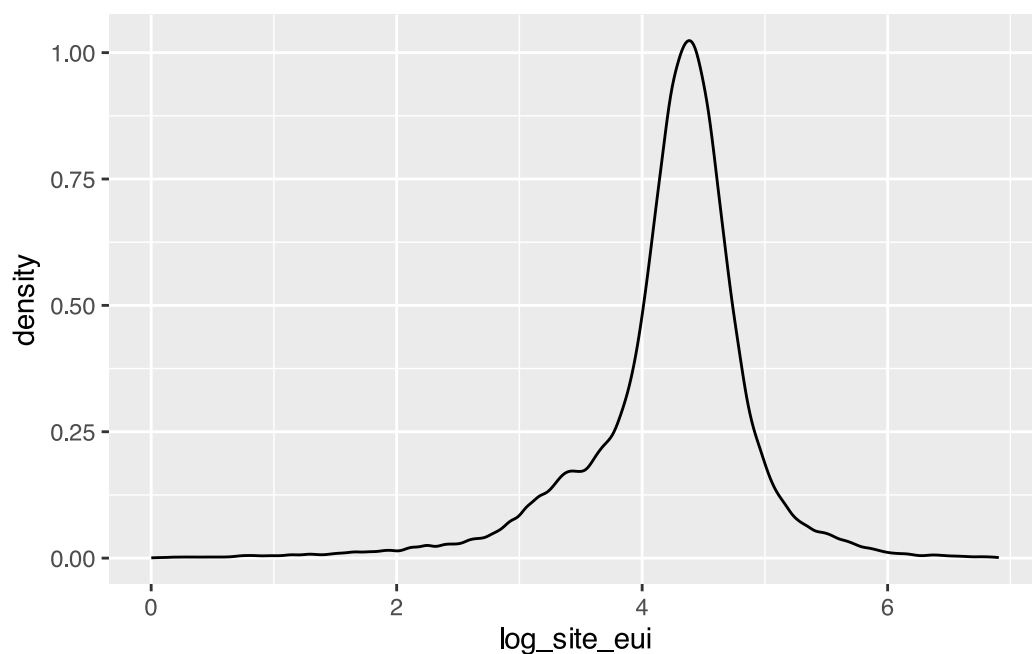
```
data <- data |>
  janitor::clean_names() |>
  mutate(
    state_factor = factor(state_factor),
    building_class = factor(building_class),
    facility_type = factor(facility_type),
    year_built = factor(year_built, ordered = TRUE),
    energy_star_rating = factor(energy_star_rating, ordered = TRUE),
    log_site_eui = log(site_eui)
  )

data |>
  gt_preview() |>
  tab_header(title = "Building Energy Dataset")
```

## Building Energy Dataset

year_	state_factor	building_class	facility_type	year_built	energy_star_rating	log_site_eui
2016	CA	Commercial	Office	2016	1	1.5
2017	CA	Commercial	Office	2017	2	1.6
2018	CA	Commercial	Office	2018	3	1.7
2019	CA	Commercial	Office	2019	4	1.8
2020	CA	Commercial	Office	2020	5	1.9
2021	CA	Commercial	Office	2021	6	2.0
2022	CA	Commercial	Office	2022	7	2.1
2023	CA	Commercial	Office	2023	8	2.2
2024	CA	Commercial	Office	2024	9	2.3
2025	CA	Commercial	Office	2025	10	2.4
2026	CA	Commercial	Office	2026	11	2.5
2027	CA	Commercial	Office	2027	12	2.6
2028	CA	Commercial	Office	2028	13	2.7
2029	CA	Commercial	Office	2029	14	2.8
2030	CA	Commercial	Office	2030	15	2.9
2031	CA	Commercial	Office	2031	16	3.0
2032	CA	Commercial	Office	2032	17	3.1
2033	CA	Commercial	Office	2033	18	3.2
2034	CA	Commercial	Office	2034	19	3.3
2035	CA	Commercial	Office	2035	20	3.4
2036	CA	Commercial	Office	2036	21	3.5
2037	CA	Commercial	Office	2037	22	3.6
2038	CA	Commercial	Office	2038	23	3.7
2039	CA	Commercial	Office	2039	24	3.8
2040	CA	Commercial	Office	2040	25	3.9
2041	CA	Commercial	Office	2041	26	4.0
2042	CA	Commercial	Office	2042	27	4.1
2043	CA	Commercial	Office	2043	28	4.2
2044	CA	Commercial	Office	2044	29	4.3
2045	CA	Commercial	Office	2045	30	4.4
2046	CA	Commercial	Office	2046	31	4.5
2047	CA	Commercial	Office	2047	32	4.6
2048	CA	Commercial	Office	2048	33	4.7
2049	CA	Commercial	Office	2049	34	4.8
2050	CA	Commercial	Office	2050	35	4.9
2051	CA	Commercial	Office	2051	36	5.0
2052	CA	Commercial	Office	2052	37	5.1
2053	CA	Commercial	Office	2053	38	5.2
2054	CA	Commercial	Office	2054	39	5.3
2055	CA	Commercial	Office	2055	40	5.4
2056	CA	Commercial	Office	2056	41	5.5
2057	CA	Commercial	Office	2057	42	5.6
2058	CA	Commercial	Office	2058	43	5.7
2059	CA	Commercial	Office	2059	44	5.8
2060	CA	Commercial	Office	2060	45	5.9
2061	CA	Commercial	Office	2061	46	6.0
2062	CA	Commercial	Office	2062	47	6.1
2063	CA	Commercial	Office	2063	48	6.2
2064	CA	Commercial	Office	2064	49	6.3
2065	CA	Commercial	Office	2065	50	6.4
2066	CA	Commercial	Office	2066	51	6.5
2067	CA	Commercial	Office	2067	52	6.6
2068	CA	Commercial	Office	2068	53	6.7
2069	CA	Commercial	Office	2069	54	6.8
2070	CA	Commercial	Office	2070	55	6.9
2071	CA	Commercial	Office	2071	56	7.0
2072	CA	Commercial	Office	2072	57	7.1
2073	CA	Commercial	Office	2073	58	7.2
2074	CA	Commercial	Office	2074	59	7.3
2075	CA	Commercial	Office	2075	60	7.4
2076	CA	Commercial	Office	2076	61	7.5
2077	CA	Commercial	Office	2077	62	7.6
2078	CA	Commercial	Office	2078	63	7.7
2079	CA	Commercial	Office	2079	64	7.8
2080	CA	Commercial	Office	2080	65	7.9
2081	CA	Commercial	Office	2081	66	8.0
2082	CA	Commercial	Office	2082	67	8.1
2083	CA	Commercial	Office	2083	68	8.2
2084	CA	Commercial	Office	2084	69	8.3
2085	CA	Commercial	Office	2085	70	8.4
2086	CA	Commercial	Office	2086	71	8.5
2087	CA	Commercial	Office	2087	72	8.6
2088	CA	Commercial	Office	2088	73	8.7
2089	CA	Commercial	Office	2089	74	8.8
2090	CA	Commercial	Office	2090	75	8.9
2091	CA	Commercial	Office	2091	76	9.0
2092	CA	Commercial	Office	2092	77	9.1
2093	CA	Commercial	Office	2093	78	9.2
2094	CA	Commercial	Office	2094	79	9.3
2095	CA	Commercial	Office	2095	80	9.4
2096	CA	Commercial	Office	2096	81	9.5
2097	CA	Commercial	Office	2097	82	9.6
2098	CA	Commercial	Office	2098	83	9.7
2099	CA	Commercial	Office	2099	84	9.8
2100	CA	Commercial	Office	2100	85	9.9
2101	CA	Commercial	Office	2101	86	10.0
2102	CA	Commercial	Office	2102	87	10.1
2103	CA	Commercial	Office	2103	88	10.2
2104	CA	Commercial	Office	2104	89	10.3
2105	CA	Commercial	Office	2105	90	10.4
2106	CA	Commercial	Office	2106	91	10.5
2107	CA	Commercial	Office	2107	92	10.6
2108	CA	Commercial	Office	2108	93	10.7
2109	CA	Commercial	Office	2109	94	10.8
2110	CA	Commercial	Office	2110	95	10.9
2111	CA	Commercial	Office	2111	96	11.0
2112	CA	Commercial	Office	2112	97	11.1
2113	CA	Commercial	Office	2113	98	11.2
2114	CA	Commercial	Office	2114	99	11.3
2115	CA	Commercial	Office	2115	100	11.4
2116	CA	Commercial	Office	2116	101	11.5
2117	CA	Commercial	Office	2117	102	11.6
2118	CA	Commercial	Office	2118	103	11.7
2119	CA	Commercial	Office	2119	104	11.8
2120	CA	Commercial	Office	2120	105	11.9
2121	CA	Commercial	Office	2121	106	12.0
2122	CA	Commercial	Office	2122	107	12.1
2123	CA	Commercial	Office	2123	108	12.2
2124	CA	Commercial	Office	2124	109	12.3
2125	CA	Commercial	Office	2125	110	12.4
2126	CA	Commercial	Office	2126	111	12.5
2127	CA	Commercial	Office	2127	112	12.6
2128	CA	Commercial	Office	2128	113	12.7
2129	CA	Commercial	Office	2129	114	12.8
2130	CA	Commercial	Office	2130	115	12.9
2131	CA	Commercial	Office	2131	116	13.0
2132	CA	Commercial	Office	2132	117	13.1
2133	CA	Commercial	Office	2133	118	13.2
2134	CA	Commercial	Office	2134	119	13.3
2135	CA	Commercial	Office	2135	120	13.4
2136	CA	Commercial	Office	2136	121	13.5
2137	CA	Commercial	Office	2137	122	13.6
2138	CA	Commercial	Office	2138	123	13.7
2139	CA	Commercial	Office	2139	124	13.8
2140	CA	Commercial	Office	2140	125	13.9
2141	CA	Commercial	Office	2141	126	14.0
2142	CA	Commercial	Office	2142	127	14.1
2143	CA	Commercial	Office	2143	128	14.2
2144	CA	Commercial	Office	2144	129	14.3
2145	CA	Commercial	Office	2145	130	14.4
2146	CA	Commercial	Office	2146	131	14.5
2147	CA	Commercial	Office	2147	132	14.6
2148	CA	Commercial	Office	2148	133	14.7
2149	CA	Commercial	Office	2149	134	14.8
2150	CA	Commercial	Office	2150	135	14.9
2151	CA	Commercial	Office	2151	136	15.0
2152	CA	Commercial	Office	2152	137	15.1
2153	CA	Commercial	Office	2153	138	15.2
2154	CA	Commercial	Office	2154	139	15.3
2155	CA	Commercial	Office	2155	140	15.4
2156	CA	Commercial	Office	2156	141	15.5
2157	CA	Commercial	Office	2157	142	15.6
2158	CA	Commercial	Office	2158	143	15.7
2159	CA	Commercial	Office	2159	144	15.8
2160	CA	Commercial	Office	2160	145	15.9
2161	CA	Commercial	Office	2161	146	16.0
2162	CA	Commercial	Office	2162	147	16.1
2163	CA	Commercial	Office	2163	148	16.2
2164	CA	Commercial	Office	2164	149	16.3
2165	CA	Commercial	Office	2165	150	16.4
2166	CA	Commercial	Office	2166	151	16.5
2167	CA	Commercial	Office	2167	152	16.6
2168	CA	Commercial	Office	2168	153	16.7
2169	CA	Commercial	Office	2169	154	16.8
2170	CA	Commercial	Office	2170	155	16.9
2171	CA	Commercial	Office	2171	156	17.0
2172	CA	Commercial	Office	2172	157	17.1
2173	CA	Commercial	Office	2173	158	17.2
2174	CA	Commercial	Office	2174	159	17.3
2175	CA	Commercial	Office	2175	160	17.4
2176	CA	Commercial	Office	2176	161	17.5
2177	CA	Commercial	Office	2177	162	17.6
2178	CA	Commercial	Office	2178	163	17.7
2179	CA	Commercial	Office	2179	164	17.8
2180	CA	Commercial	Office	2180	165	17.9
2181	CA	Commercial	Office	2181	166	18.0
2182	CA	Commercial	Office	2182	167	18.1
2183	CA	Commercial	Office	2183	168	18.2
2184	CA	Commercial	Office	2184	169	18.3
2185	CA	Commercial	Office	2185	170	18.4
2186	CA	Commercial	Office	2186	171	18.5
2187	CA	Commercial	Office	2187	172	18.6
2188	CA	Commercial	Office	2188	173	18.7
2189	CA	Commercial	Office	2189	174	18.8
2190	CA	Commercial	Office	2190	175	18.9
2191	CA	Commercial	Office	2191	176	19.0
2192	CA	Commercial	Office	2192	177	19.1
2193	CA	Commercial	Office	2193	178	19.2
2194	CA	Commercial	Office	2194	179	19.3
2195	CA	Commercial	Office	2195	180	19.4
2196	CA	Commercial	Office	2196	181	19.5
2197	CA	Commercial	Office	2197	182	19.6
2198	CA	Commercial	Office	2198	183	19.7
2199	CA	Commercial	Office	2199	184	19.8
2200	CA	Commercial	Office	2200	185	19.9
2201	CA	Commercial	Office	2201	186	20.0
2202	CA	Commercial	Office	2202	187	20.1
2203	CA	Commercial	Office	2203	188	20.2
2204	CA	Commercial	Office	2204	189	20.3
2205	CA	Commercial	Office	2205	190	20.4
2206	CA	Commercial	Office	2206	191	20.5
2207	CA	Commercial	Office	2207	192	20.6
2208	CA	Commercial	Office	2208	193	20.7
2209	CA	Commercial	Office	2209	194	20.8
2210	CA	Commercial	Office	2210	195	20.9
2211	CA	Commercial	Office	2211	196	21.0
2212	CA	Commercial	Office	2212	197	21.1
2213	CA	Commercial	Office	2213	198	21.2
2214	CA	Commercial	Office	2214	199	21.3
2215	CA	Commercial	Office	2215	200	21.4
2216	CA	Commercial	Office	2216	201	21.5
2217	CA	Commercial	Office	2217	202	21.6
2218	CA	Commercial	Office	2218	203	21.7
2219	CA	Commercial	Office	2219	204	21.8
2220	CA	Commercial	Office	2220	205	21.9
2221	CA	Commercial	Office	2221	206	22.0
2222	CA	Commercial	Office	2222	207	22.1
2223	CA	Commercial	Office	2223	208	22.2
222						

```
data |>
  ggplot(mapping = aes(x = log_site_eui)) +
  geom_density()
```



```
res <- lm(site_eui ~ facility_type, data = data)
summary(res)
```

Call:

```
lm(formula = site_eui ~ facility_type, data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-317.99	-22.06	-5.52	13.75	909.55

Coefficients:

	Estimate	Std. Error
(Intercept)	31.877	1.182
facility_type5plus_Unit_Building	4.861	1.864
facility_typeCommercial_Other	60.765	1.707
facility_typeCommercial_Unknown	81.275	5.406
facility_typeData_Center	307.858	9.965
facility_typeEducation_College_or_university	76.752	1.975
facility_typeEducation_Other_classroom	37.565	1.443
facility_typeEducation_Preschool_or_daycare	29.097	5.087
facility_typeEducation_Uncategorized	14.296	2.264
facility_typeFood_Sales	104.916	6.258
facility_typeFood_Service_Other	-2.640	12.526
facility_typeFood_Service_Restaurant_or_cafeteria	163.717	6.535
facility_typeFood_Service_Uncategorized	96.925	12.176

facility_typeGrocery_store_or_food_market	209.258	2.701
facility_typeHealth_Care_Inpatient	216.464	2.804
facility_typeHealth_Care_Outpatient_Clinic	71.736	6.972
facility_typeHealth_Care_Outpatient_Uncategorized	158.015	8.650
facility_typeHealth_Care_Uncategorized	152.068	7.296
facility_typeIndustrial	93.468	2.878
facility_typeLaboratory	297.572	5.109
facility_typeLodging_Dormitory_or_fraternity_sorority	49.719	2.313
facility_typeLodging_Hotel	73.058	1.630
facility_typeLodging_Other	89.081	6.053
facility_typeLodging_Uncategorized	34.719	23.024
facility_typeMixed_Use_Commercial_and_Residential	57.653	2.309
facility_typeMixed_Use_Predominantly_Commercial	37.256	3.424
facility_typeMixed_Use_Predominantly_Residential	49.921	17.179
facility_typeMultifamily_Uncategorized	52.002	1.210
facility_typeNursing_Home	99.437	2.196
facility_typeOffice_Bank_or_other_financial	58.019	4.084
facility_typeOffice_Medical_non_diagnostic	84.885	2.704
facility_typeOffice_Mixed_use	50.233	12.176
facility_typeOffice_Uncategorized	45.197	1.268
facility_typeParking_Garage	35.474	3.454
facility_typePublic_Assembly_Drama_theater	49.040	6.258
facility_typePublic_Assembly_Entertainment_culture	87.023	5.043
facility_typePublic_Assembly_Library	73.972	4.233
facility_typePublic_Assembly_Movie_Theater	71.218	8.317
facility_typePublic_Assembly_Other	94.827	4.474
facility_typePublic_Assembly_Recreation	83.301	6.174
facility_typePublic_Assembly_Social_meeting	47.045	5.607
facility_typePublic_Assembly_Stadium	125.165	17.179
facility_typePublic_Assembly_Uncategorized	30.996	10.351
facility_typePublic_Safety_Courthouse	71.305	8.424
facility_typePublic_Safety_Fire_or_police_station	99.246	4.270
facility_typePublic_Safety_Penitentiary	139.028	8.535
facility_typePublic_Safety_Uncategorized	51.832	7.929
facility_typeReligious_worship	12.684	2.832
facility_typeRetail_Enclosed_mall	69.089	4.840
facility_typeRetail_Strip_shopping_mall	78.542	4.979
facility_typeRetail_Uncategorized	49.026	1.933
facility_typeRetail_Vehicle_dealership_showroom	14.764	6.093
facility_typeService_Drycleaning_or_Laundry	10.236	17.179
facility_typeService_Uncategorized	81.689	6.346
facility_typeService_Vehicle_service_repair_shop	105.719	4.533
facility_typeWarehouse_Distribution_or_Shipping_center	7.683	2.403
facility_typeWarehouse_Nonrefrigerated	6.332	1.872
facility_typeWarehouse_Refrigerated	64.648	4.979
facility_typeWarehouse_Selfstorage	-10.288	2.445
facility_typeWarehouse_Uncategorized	4.060	3.067
t value Pr(> t )		

(Intercept)	26.975	< 2e-16	***
facility_type5plus_Unit_Building	2.608	0.009101	**
facility_typeCommercial_Other	35.607	< 2e-16	***
facility_typeCommercial_Unknown	15.035	< 2e-16	***
facility_typeData_Center	30.893	< 2e-16	***
facility_typeEducation_College_or_university	38.866	< 2e-16	***
facility_typeEducation_Other_classroom	26.038	< 2e-16	***
facility_typeEducation_Preschool_or_daycare	5.720	1.07e-08	***
facility_typeEducation_Uncategorized	6.315	2.72e-10	***
facility_typeFood_Sales	16.765	< 2e-16	***
facility_typeFood_Service_Other	-0.211	0.833054	
facility_typeFood_Service_Restaurant_or_cafeteria	25.054	< 2e-16	***
facility_typeFood_Service_Uncategorized	7.960	1.74e-15	***
facility_typeGrocery_store_or_food_market	77.465	< 2e-16	***
facility_typeHealth_Care_Inpatient	77.211	< 2e-16	***
facility_typeHealth_Care_Outpatient_Clinic	10.290	< 2e-16	***
facility_typeHealth_Care_Outpatient_Uncategorized	18.267	< 2e-16	***
facility_typeHealth_Care_Uncategorized	20.843	< 2e-16	***
facility_typeIndustrial	32.481	< 2e-16	***
facility_typeLaboratory	58.244	< 2e-16	***
facility_typeLodging_Dormitory_or_fraternity_sorority	21.499	< 2e-16	***
facility_typeLodging_Hotel	44.824	< 2e-16	***
facility_typeLodging_Other	14.716	< 2e-16	***
facility_typeLodging_Uncategorized	1.508	0.131577	
facility_typeMixed_Use_Commercial_and_Residential	24.971	< 2e-16	***
facility_typeMixed_Use_Predominantly_Commercial	10.881	< 2e-16	***
facility_typeMixed_Use_Predominantly_Residential	2.906	0.003663	**
facility_typeMultifamily_Uncategorized	42.986	< 2e-16	***
facility_typeNursing_Home	45.289	< 2e-16	***
facility_typeOffice_Bank_or_other_financial	14.207	< 2e-16	***
facility_typeOffice_Medical_non_diagnostic	31.395	< 2e-16	***
facility_typeOffice_Mixed_use	4.126	3.70e-05	***
facility_typeOffice_Uncategorized	35.645	< 2e-16	***
facility_typeParking_Garage	10.271	< 2e-16	***
facility_typePublic_Assembly_Drama_theater	7.836	4.69e-15	***
facility_typePublic_Assembly_Entertainment_culture	17.257	< 2e-16	***
facility_typePublic_Assembly_Library	17.475	< 2e-16	***
facility_typePublic_Assembly_Movie_Theater	8.563	< 2e-16	***
facility_typePublic_Assembly_Other	21.197	< 2e-16	***
facility_typePublic_Assembly_Recreation	13.493	< 2e-16	***
facility_typePublic_Assembly_Social_meeting	8.391	< 2e-16	***
facility_typePublic_Assembly_Stadium	7.286	3.23e-13	***
facility_typePublic_Assembly_Uncategorized	2.995	0.002750	**
facility_typePublic_Safety_Courthouse	8.465	< 2e-16	***
facility_typePublic_Safety_Fire_or_police_station	23.242	< 2e-16	***
facility_typePublic_Safety_Penitentiary	16.289	< 2e-16	***
facility_typePublic_Safety_Uncategorized	6.537	6.33e-11	***
facility_typeReligious_worship	4.478	7.54e-06	***

facility_typeRetail_Enclosed_mall	14.274	< 2e-16	***
facility_typeRetail_Strip_shopping_mall	15.775	< 2e-16	***
facility_typeRetail_Uncategorized	25.365	< 2e-16	***
facility_typeRetail_Vehicle_dealership_showroom	2.423	0.015384	*
facility_typeService_Drycleaning_or_Laundry	0.596	0.551299	
facility_typeService_Uncategorized	12.872	< 2e-16	***
facility_typeService_Vehicle_service_repair_shop	23.320	< 2e-16	***
facility_typeWarehouse_Distribution_or_Shipping_center	3.197	0.001387	**
facility_typeWarehouse_Nonrefrigerated	3.383	0.000716	***
facility_typeWarehouse_Refrigerated	12.984	< 2e-16	***
facility_typeWarehouse_Selfstorage	-4.208	2.58e-05	***
facility_typeWarehouse_Uncategorized	1.324	0.185652	

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 51.42 on 75697 degrees of freedom

Multiple R-squared: 0.2217, Adjusted R-squared: 0.221

F-statistic: 365.4 on 59 and 75697 DF, p-value: < 2.2e-16

## Bibliography