Health Indicators to Combat Obesity, Heart Disease and Cancer

Milestone 7: FINAL REPORT
Cloud Project
Cloud chosen: Google Cloud Platform

GROUP 04
Aditi Namdeo
Arjun Janardhan

namdeo.a@northeastern.edu
janardhan.a@northeastern.edu

Percentage of effort by Aditi: 50%
Percentage of effort by Arjun: 50%

Signature of Student 01: Aditi Namdeo
Signature of Student 02: Arjun Janardhan
Submission Date: 31st March, 202

PROBLEM SETTING

Obesity increases the risk of several debilitating, and deadly diseases, including diabetes, heart disease, and some cancers. It does this through a variety of pathways, some as straightforward as the mechanical stress of carrying extra pounds and some involving complex changes in hormones and metabolism. There are many reasons why some people have difficulty losing weight. Usually, obesity results from inherited, physiological and environmental factors, combined with diet, physical activity and exercise choices. In this project, Community Health Status Indicators (CHSI) to combat obesity, heart disease, and cancer are major components of the Community Health Data Initiative. The selected dataset provides key health indicators for local communities and encourages dialogue about actions that can be taken to improve community health (e.g., obesity, heart disease, cancer). The health indicators are an important discussion to empower health consciousness and spread awareness about the ill effects of obesity and factors that cause the same.

PROBLEM DEFINITION

Community Health Status Indicators (CHSI) to combat obesity, heart disease, and cancer are major components of the Community Health Data Initiative. The dataset provides key health indicators for local communities and encourages dialogue about actions that can be taken to improve community health (e.g., obesity, heart disease, cancer). The CHSI report and dataset was designed not only for public health professionals but also for members of the community who are interested in the health of their community. The CHSI report contains over 200 measures for each of the 3,141 United States counties. Although CHSI presents indicators like deaths due to heart disease and cancer, it is imperative to understand that behavioral factors such as obesity, tobacco use, diet, physical activity, alcohol and drug use, sexual behavior and others substantially contribute to these deaths.

Our team is challenged to undertake research or analysis on this data and submit the findings. This project's purpose is to use data engineering and warehousing concepts to build data pipelines that receive data from a source, transform it, and store it in the best possible format for data visualization and to derive actionable and scalable insights from the data. We are trying to answer the following questions:

- What are the major factors leading to obesity, heart diseases and cancer?
- What is the reason behind largest number of deaths?
- Top few factors of health illness in people?
- What are some ways to improve mortality rate due to these health conditions?

DATA SOURCES

Community Health Status Indicators (CHSI) to combat obesity, heart disease, and cancer are major components of the Community Health Data Initiative. This dataset provides key health indicators for local communities and encourages dialogue about actions that can be taken to improve community health (e.g., obesity, heart disease, cancer). The CHSI report and dataset was designed not only for public health professionals but also for members of the community who are interested in the health of their community. The CHSI report contains over 200 measures for each of the 3,141 United States counties. Although CHSI presents indicators like deaths due to heart disease and cancer, it is imperative to understand that behavioral factors such as obesity, tobacco use, diet, physical activity, alcohol and drug use, sexual behavior and others substantially contribute to these deaths.

Citation-

Source: https://catalog.data.gov/dataset/community-health-status-indicators-chsi-to-combat-obesity-heart-disease-and-cancer

DATA DESCRIPTION – ANALYSIS DIMENSIONS

We have 9 different health datasets and 3 data defining datasets:

All datasets in total have 1180 entity columns.

Following is the description of each dataset:

DATA_ELEMENT_DESCRIPTION.csv defines each data element and indicates where its description is found in Data Sources, Definitions, and Notes.

DEFINED_DATA_VALUE.csv defines the meaning of specific values (such as missing or suppressed data).

HEALTHY_PEOPLE_2010.csv identifies the Healthy People 2010 Targets and the U.S. Percentages or Rates.

DEMOGRAPHICS.csv identifies the data elements and values in the Demographics indicator domain.

LEADING_CAUSES_OF_DEATH.csv identifies the data elements and values in the Leading Causes of Death indicator domain.

SUMMARY_MEASURES_OF_HEALTH.csv identifies the data elements and values in the Summary Measures of Health indicator domain.

MEASURES_OF_BIRTH_AND_DEATH.csv identifies the data elements and values in the Measures of Birth and Death indicator domain.

RELATIVE_HEALTH_IMPORTANCE.csv identifies the data elements and values in the Relative Health Importance indicator domain.

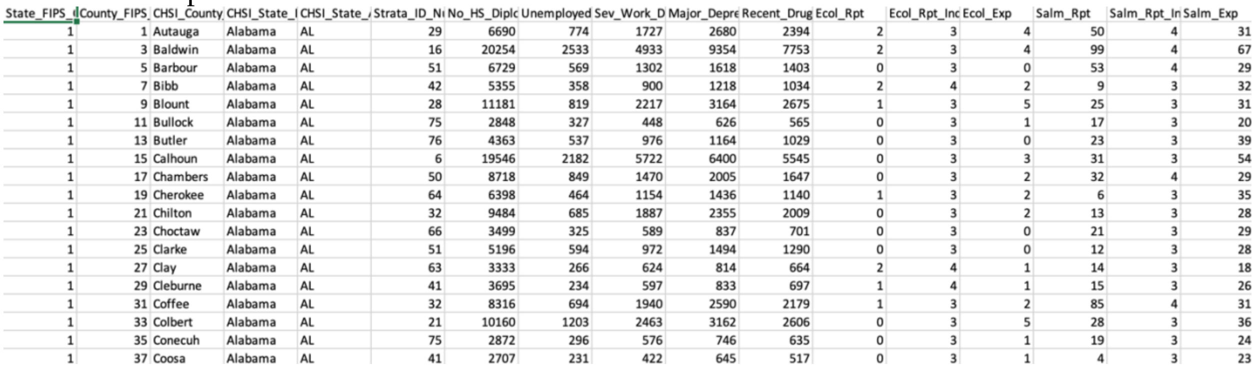
VULNERABLE_POPS_AND_ENV_HEALTH.csv identifies the data elements and values in the Vulnerable Populations and Environmental Health indicator domain.

PREVENTIVE_SERVICES_USE.csv identifies the data elements and values in the Preventive Services indicator domain.

RISK_FACTORS_AND_ACCESS_TO_CARE.csv identifies the data elements and values in the Risk Factors and Access to Care indicator domain.

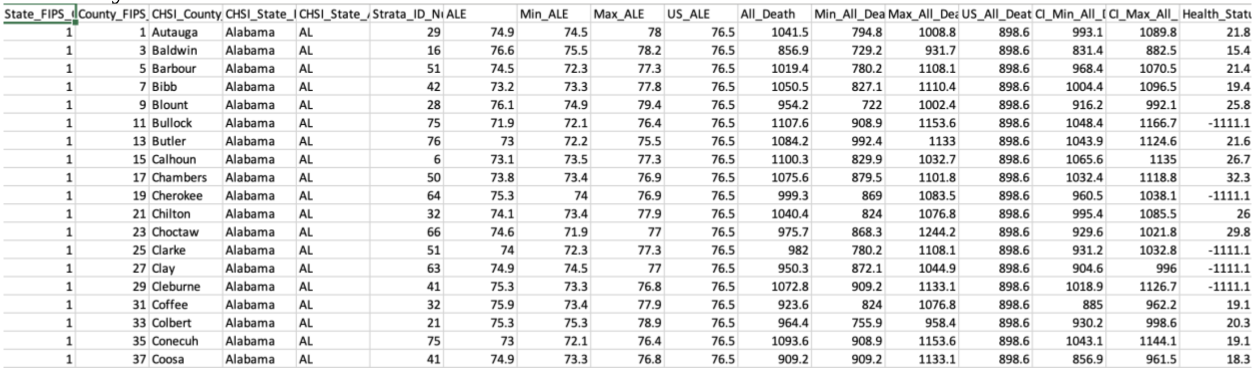
Following are the headers of each dataset’s CSV:

Vulnerable Population and Environment Health CSV



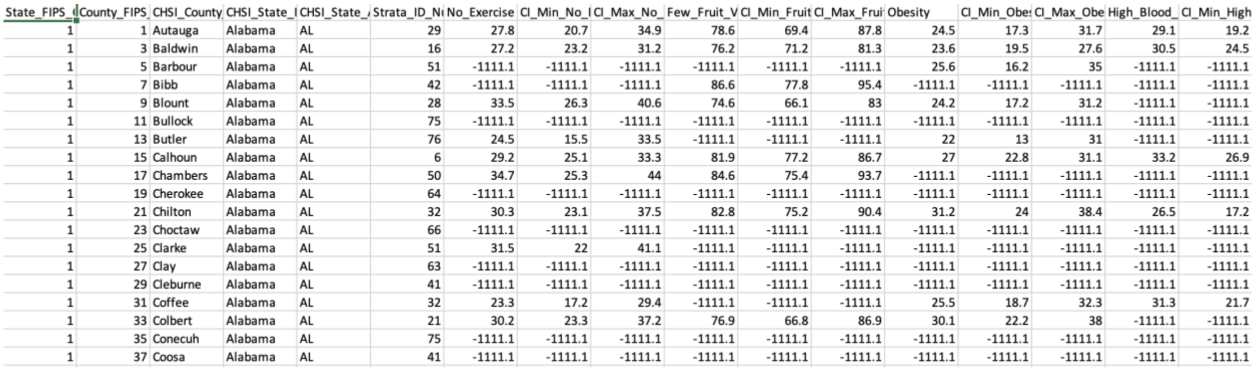
State_FIPS_	County_FIPS_	CHSI_County	CHSI_State_	CHSI_State_	Strata_ID_N	No_HS_Diplc	Unemployed	Sev_Work_D	Major_Depre	Recent_Drug	Ecol_Rpt	Ecol_Rpt_Inc	Ecol_Exp	Salm_Rpt	Salm_Rpt_In	Salm_Exp
1	1	Autauga	Alabama	AL	29	6690	774	1727	2680	2394	2	3	4	50	4	31
1	3	Baldwin	Alabama	AL	16	20254	2533	4933	9354	7753	2	3	4	99	4	67
1	5	Barbour	Alabama	AL	51	6729	569	1302	1618	1403	0	3	0	53	4	29
1	7	Bibb	Alabama	AL	42	5355	358	900	1218	1034	2	4	2	9	3	32
1	9	Blount	Alabama	AL	28	11181	819	2217	3164	2675	1	3	5	25	3	31
1	11	Bullock	Alabama	AL	75	2848	327	448	626	565	0	3	1	17	3	20
1	13	Butler	Alabama	AL	76	4363	537	976	1164	1029	0	3	0	23	3	39
1	15	Calhoun	Alabama	AL	6	19546	2182	5722	6400	5545	0	3	3	31	3	54
1	17	Chambers	Alabama	AL	50	8718	849	1470	2005	1647	0	3	2	32	4	29
1	19	Cherokee	Alabama	AL	64	6398	464	1154	1436	1140	1	3	2	6	3	35
1	21	Chilton	Alabama	AL	32	9484	685	1887	2355	2009	0	3	2	13	3	28
1	23	Choctaw	Alabama	AL	66	3499	325	589	837	701	0	3	0	21	3	29
1	25	Clarke	Alabama	AL	51	5196	594	972	1494	1290	0	3	0	12	3	28
1	27	Clay	Alabama	AL	63	3333	266	624	814	664	2	4	1	14	3	18
1	29	Cleburne	Alabama	AL	41	3695	234	597	833	697	1	4	1	15	3	26
1	31	Coffee	Alabama	AL	32	8316	694	1940	2590	2179	1	3	2	85	4	31
1	33	Colbert	Alabama	AL	21	10160	1203	2463	3162	2606	0	3	5	28	3	36
1	35	Conecuh	Alabama	AL	75	2872	296	576	746	635	0	3	1	19	3	24
1	37	Coosa	Alabama	AL	41	2707	231	422	645	517	0	3	1	4	3	23

Summary Measures of Health CSV



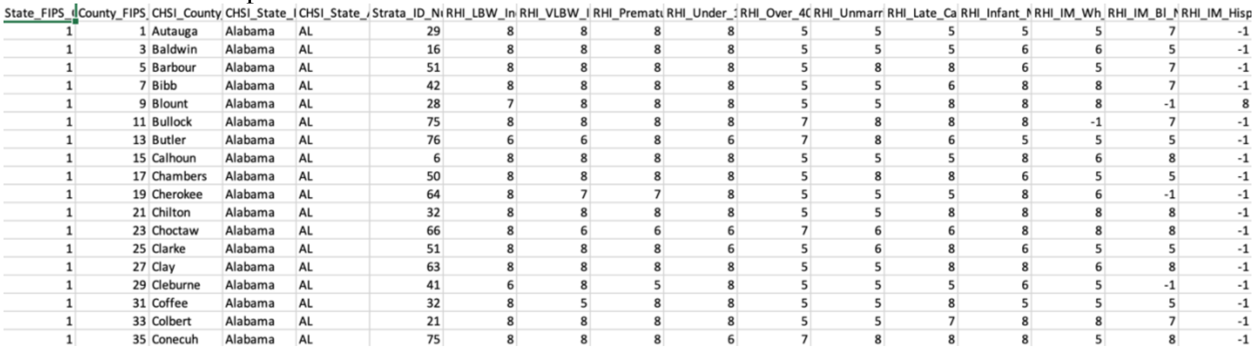
State_FIPS_	County_FIPS_	CHSI_County	CHSI_State_	CHSI_State_	Strata_ID_N	ALE	Min_ALE	Max_ALE	US_ALE	All_Death	Min_All_Dea	Max_All_Dea	US_All_Deat	CI_Min_All_	CI_Max_All_	Health_Statu
1	1	Autauga	Alabama	AL	29	74.9	74.5	78	76.5	1041.5	794.8	1008.8	898.6	993.1	1089.8	21.8
1	3	Baldwin	Alabama	AL	16	76.6	75.5	78.2	76.5	856.9	729.2	931.7	898.6	831.4	882.5	15.4
1	5	Barbour	Alabama	AL	51	74.5	72.3	77.3	76.5	1019.4	780.2	1108.1	898.6	968.4	1070.5	21.4
1	7	Bibb	Alabama	AL	42	73.2	73.3	77.8	76.5	1050.5	827.1	1110.4	898.6	1004.4	1096.5	19.4
1	9	Blount	Alabama	AL	28	76.1	74.9	79.4	76.5	954.2	722	1002.4	898.6	916.2	992.1	25.8
1	11	Bullock	Alabama	AL	75	71.9	72.1	76.4	76.5	1107.6	908.9	1153.6	898.6	1048.4	1166.7	-1111.1
1	13	Butler	Alabama	AL	76	73	72.2	75.5	76.5	1084.2	992.4	1133	898.6	1043.9	1124.6	21.6
1	15	Calhoun	Alabama	AL	6	73.1	73.5	77.3	76.5	1100.3	829.9	1032.7	898.6	1065.6	1135	26.7
1	17	Chambers	Alabama	AL	50	73.8	73.4	76.9	76.5	1075.6	879.5	1101.8	898.6	1032.4	1118.8	32.3
1	19	Cherokee	Alabama	AL	64	75.3	74	76.9	76.5	999.3	869	1083.5	898.6	960.5	1038.1	-1111.1
1	21	Chilton	Alabama	AL	32	74.1	73.4	77.9	76.5	1040.4	824	1076.8	898.6	995.4	1085.5	26
1	23	Choctaw	Alabama	AL	66	74.6	71.9	77	76.5	975.7	868.3	1244.2	898.6	929.6	1021.8	29.8
1	25	Clarke	Alabama	AL	51	74	72.3	77.3	76.5	982	780.2	1108.1	898.6	931.2	1032.8	-1111.1
1	27	Clay	Alabama	AL	63	74.9	74.5	77	76.5	950.3	872.1	1044.9	898.6	904.6	996	-1111.1
1	29	Cleburne	Alabama	AL	41	75.3	73.3	76.8	76.5	1072.8	909.2	1133.1	898.6	1018.9	1126.7	-1111.1
1	31	Coffee	Alabama	AL	32	75.9	73.4	77.9	76.5	923.6	824	1076.8	898.6	885	962.2	19.1
1	33	Colbert	Alabama	AL	21	75.3	75.3	78.9	76.5	964.4	755.9	958.4	898.6	930.2	998.6	20.3
1	35	Conecuh	Alabama	AL	75	73	72.1	76.4	76.5	1093.6	908.9	1153.6	898.6	1043.1	1144.1	19.1
1	37	Coosa	Alabama	AL	41	74.9	73.3	76.8	76.5	909.2	909.2	1133.1	898.6	856.9	961.5	18.3

Risk Factors and Access to Healthcare CSV



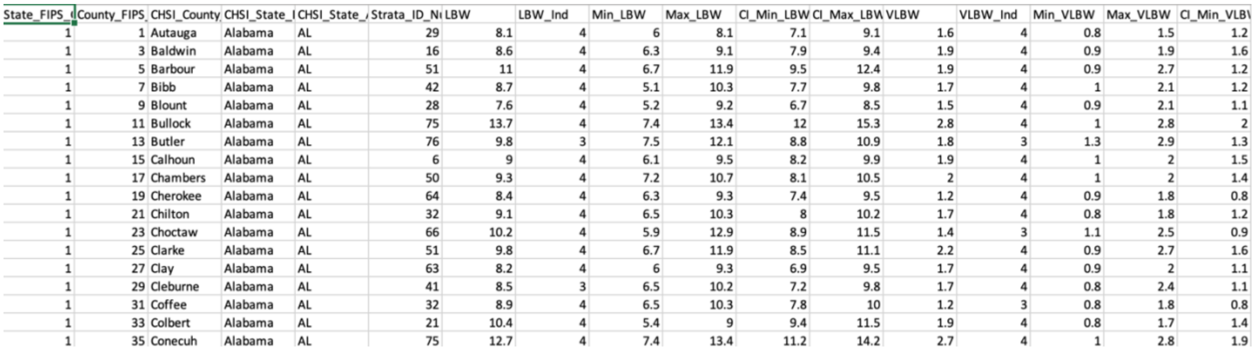
State_FIPS_	County_FIPS_	CHSI_County	CHSI_State_	CHSI_State_	Strata_ID_N	No_Exercise	CI_Min_No_	CI_Max_No_	Few_Fruit_V	CI_Min_Fruit	CI_Max_Frui	Obesity	CI_Min_Obe	CI_Max_Obe	High_Blood_	CI_Min_High
1	1	Autauga	Alabama	AL	29	27.8	20.7	34.9	78.6	69.4	87.8	24.5	17.3	31.7	29.1	19.2
1	3	Baldwin	Alabama	AL	16	27.2	23.2	31.2	76.2	71.2	81.3	23.6	19.5	27.6	30.5	24.5
1	5	Barbour	Alabama	AL	51	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	25.6	16.2	35	-1111.1	-1111.1
1	7	Bibb	Alabama	AL	42	-1111.1	-1111.1	-1111.1	86.6	77.8	95.4	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1
1	9	Blount	Alabama	AL	28	33.5	26.3	40.6	74.6	66.1	83	24.2	17.2	31.2	-1111.1	-1111.1
1	11	Bullock	Alabama	AL	75	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1
1	13	Butler	Alabama	AL	76	24.5	15.5	33.5	-1111.1	-1111.1	-1111.1	22	13	31	-1111.1	-1111.1
1	15	Calhoun	Alabama	AL	6	29.2	25.1	33.3	81.9	77.2	86.7	27	22.8	31.1	33.2	26.9
1	17	Chambers	Alabama	AL	50	34.7	25.3	44	84.6	75.4	93.7	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1
1	19	Cherokee	Alabama	AL	64	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1
1	21	Chilton	Alabama	AL	32	30.3	23.1	37.5	82.8	75.2	90.4	31.2	24	38.4	26.5	17.2
1	23	Choctaw	Alabama	AL	66	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1
1	25	Clarke	Alabama	AL	51	31.5	22	41.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1
1	27	Clay	Alabama	AL	63	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1
1	29	Cleburne	Alabama	AL	41	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1
1	31	Coffee	Alabama	AL	32	23.3	17.2	29.4	-1111.1	-1111.1	-1111.1	25.5	18.7	32.3	31.3	21.7
1	33	Colbert	Alabama	AL	21	30.2	23.3	37.2	76.9	66.8	86.9	30.1	22.2	38	-1111.1	-1111.1
1	35	Conecuh	Alabama	AL	75	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1
1	37	Coosa	Alabama	AL	41	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1	-1111.1

Relative Health Importance CSV



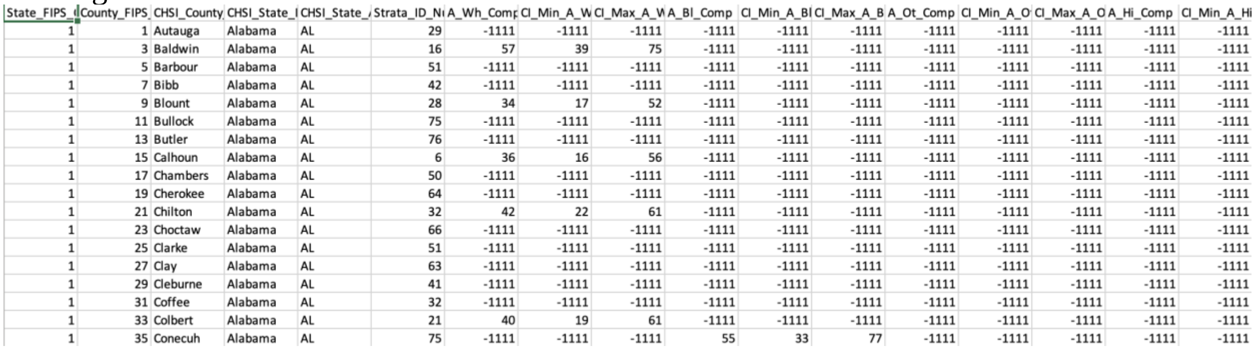
State_FIPS_(	County_FIPS_	CHSI_County	CHSI_State_I	CHSI_State_/	Strata_ID_N	RHI_LBW_In	RHI_VLBW_I	RHI_Premat	RHI_Under_1	RHI_Over_40	RHI_Unmarr	RHI_Late_Ca	RHI_Infant_N	RHI_IM_Wh_	RHI_IM_Bl_N	RHI_IM_Hisp
1	1	Autauga	Alabama	AL	29	8	8	8	8	5	5	5	5	5	7	-1
1	3	Baldwin	Alabama	AL	16	8	8	8	8	5	5	5	6	6	5	-1
1	5	Barbour	Alabama	AL	51	8	8	8	8	5	8	8	6	5	7	-1
1	7	Bibb	Alabama	AL	42	8	8	8	8	5	5	6	8	8	7	-1
1	9	Blount	Alabama	AL	28	7	8	8	8	5	5	8	8	8	-1	8
1	11	Bullock	Alabama	AL	75	8	8	8	8	7	8	8	8	-1	7	-1
1	13	Butler	Alabama	AL	76	6	6	8	6	7	8	6	5	5	5	-1
1	15	Calhoun	Alabama	AL	6	8	8	8	8	5	5	5	8	6	8	-1
1	17	Chambers	Alabama	AL	50	8	8	8	8	5	8	8	6	5	5	-1
1	19	Cherokee	Alabama	AL	64	8	7	7	8	5	5	5	8	6	-1	-1
1	21	Chilton	Alabama	AL	32	8	8	8	8	5	5	8	8	8	8	-1
1	23	Choctaw	Alabama	AL	66	8	6	6	6	7	6	6	8	8	8	-1
1	25	Clarke	Alabama	AL	51	8	8	8	6	5	6	8	6	5	5	-1
1	27	Clay	Alabama	AL	63	8	8	8	8	5	5	8	8	6	8	-1
1	29	Cleburne	Alabama	AL	41	6	8	5	8	5	5	5	6	5	-1	-1
1	31	Coffee	Alabama	AL	32	8	5	8	8	5	5	8	5	5	5	-1
1	33	Colbert	Alabama	AL	21	8	8	8	8	5	5	7	8	8	7	-1
1	35	Conecuh	Alabama	AL	75	8	8	8	6	7	8	8	8	5	8	-1

Measures of Birth and Death CSV



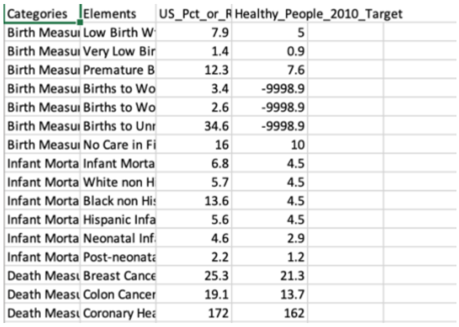
State_FIPS_(	County_FIPS_	CHSI_County	CHSI_State_I	CHSI_State_/	Strata_ID_N	LBW	LBW_Ind	Min_LBW	Max_LBW	CI_Min_LBW	CI_Max_LBW	VLBW	VLBW_Ind	Min_VLBW	Max_VLBW	CI_Min_VLBV
1	1	Autauga	Alabama	AL	29	8.1	4	6	8.1	7.1	9.1	1.6	4	0.8	1.5	1.2
1	3	Baldwin	Alabama	AL	16	8.6	4	6.3	9.1	7.9	9.4	1.9	4	0.9	1.9	1.6
1	5	Barbour	Alabama	AL	51	11	4	6.7	11.9	9.5	12.4	1.9	4	0.9	2.7	1.2
1	7	Bibb	Alabama	AL	42	8.7	4	5.1	10.3	7.7	9.8	1.7	4	1	2.1	1.2
1	9	Blount	Alabama	AL	28	7.6	4	5.2	9.2	6.7	8.5	1.5	4	0.9	2.1	1.1
1	11	Bullock	Alabama	AL	75	13.7	4	7.4	13.4	12	15.3	2.8	4	1	2.8	2
1	13	Butler	Alabama	AL	76	9.8	3	7.5	12.1	8.8	10.9	1.8	3	1.3	2.9	1.3
1	15	Calhoun	Alabama	AL	6	9	4	6.1	9.5	8.2	9.9	1.9	4	1	2	1.5
1	17	Chambers	Alabama	AL	50	9.3	4	7.2	10.7	8.1	10.5	2	4	1	2	1.4
1	19	Cherokee	Alabama	AL	64	8.4	4	6.3	9.3	7.4	9.5	1.2	4	0.9	1.8	0.8
1	21	Chilton	Alabama	AL	32	9.1	4	6.5	10.3	8	10.2	1.7	4	0.8	1.8	1.2
1	23	Choctaw	Alabama	AL	66	10.2	4	5.9	12.9	8.9	11.5	1.4	3	1.1	2.5	0.9
1	25	Clarke	Alabama	AL	51	9.8	4	6.7	11.9	8.5	11.1	2.2	4	0.9	2.7	1.6
1	27	Clay	Alabama	AL	63	8.2	4	6	9.3	6.9	9.5	1.7	4	0.9	2	1.1
1	29	Cleburne	Alabama	AL	41	8.5	3	6.5	10.2	7.2	9.8	1.7	4	0.8	2.4	1.1
1	31	Coffee	Alabama	AL	32	8.9	4	6.5	10.3	7.8	10	1.2	3	0.8	1.8	0.8
1	33	Colbert	Alabama	AL	21	10.4	4	5.4	9	9.4	11.5	1.9	4	0.8	1.7	1.4
1	35	Conecuh	Alabama	AL	75	12.7	4	7.4	13.4	11.2	14.2	2.7	4	1	2.8	1.9

Leading Causes of Death CSV



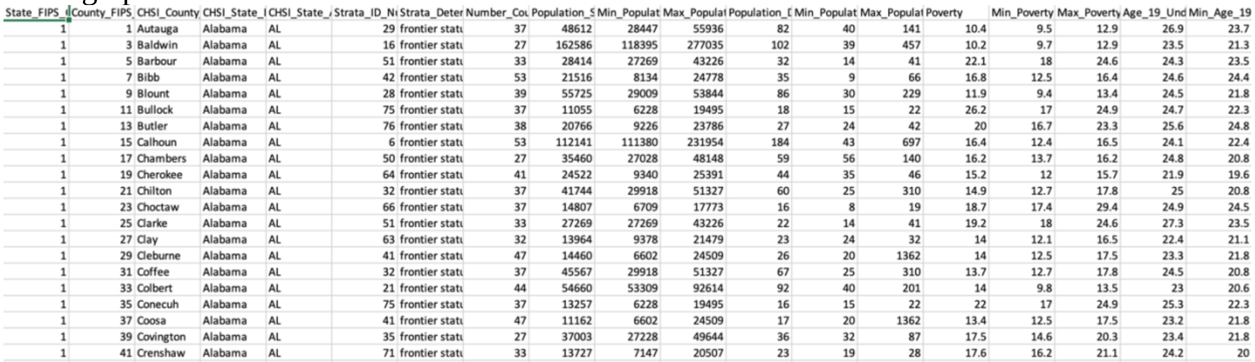
State_FIPS_(	County_FIPS_	CHSI_County	CHSI_State_I	CHSI_State_/	Strata_ID_N	A_Wh_Comp	CI_Min_A_W	CI_Max_A_W	A_Bl_Comp	CI_Min_A_Bl	CI_Max_A_B	A_Ot_Comp	CI_Min_A_O	CI_Max_A_O	A_Hi_Comp	CI_Min_A_Hi
1	1	Autauga	Alabama	AL	29	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111
1	3	Baldwin	Alabama	AL	16	57	39	75	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111
1	5	Barbour	Alabama	AL	51	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111
1	7	Bibb	Alabama	AL	42	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111
1	9	Blount	Alabama	AL	28	34	17	52	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111
1	11	Bullock	Alabama	AL	75	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111
1	13	Butler	Alabama	AL	76	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111
1	15	Calhoun	Alabama	AL	6	36	16	56	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111
1	17	Chambers	Alabama	AL	50	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111
1	19	Cherokee	Alabama	AL	64	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111
1	21	Chilton	Alabama	AL	32	42	22	61	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111
1	23	Choctaw	Alabama	AL	66	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111
1	25	Clarke	Alabama	AL	51	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111
1	27	Clay	Alabama	AL	63	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111
1	29	Cleburne	Alabama	AL	41	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111
1	31	Coffee	Alabama	AL	32	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111
1	33	Colbert	Alabama	AL	21	40	19	61	-1111	-1111	-1111	-1111	-1111	-1111	-1111	-1111
1	35	Conecuh	Alabama	AL	75	-1111	-1111	-1111	55	33	77	-1111	-1111	-1111	-1111	-1111

Healthy People 2010 CSV



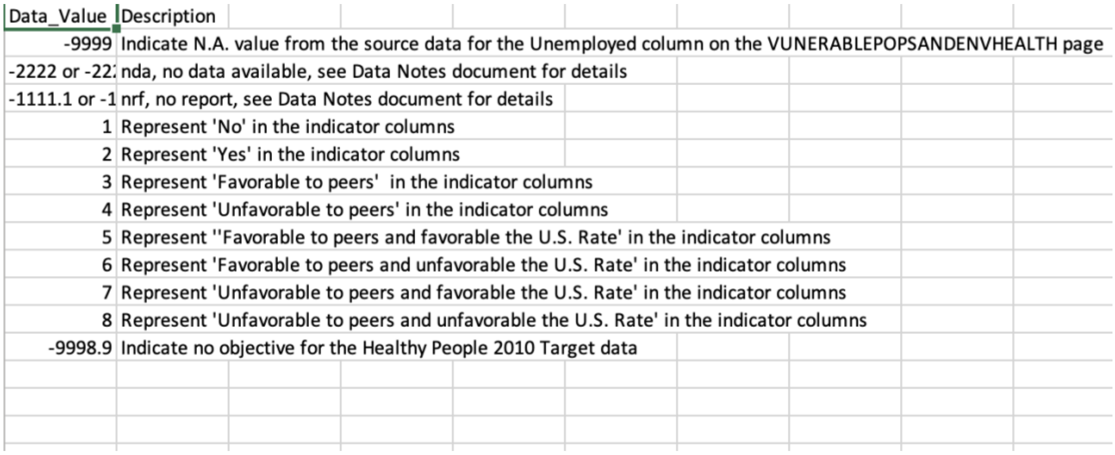
Categories	Elements	US_Pct_or_R	Healthy_People_2010_Target
Birth Measu	Low Birth W	7.9	5
Birth Measu	Very Low Bir	1.4	0.9
Birth Measu	Premature B	12.3	7.6
Birth Measu	Births to Wo	3.4	-9998.9
Birth Measu	Births to Wo	2.6	-9998.9
Birth Measu	Births to Unr	34.6	-9998.9
Birth Measu	No Care in Fi	16	10
Infant Morta	Infant Morta	6.8	4.5
Infant Morta	White non H	5.7	4.5
Infant Morta	Black non His	13.6	4.5
Infant Morta	Hispanic Infa	5.6	4.5
Infant Morta	Neonatal Inf	4.6	2.9
Infant Morta	Post-neonata	2.2	1.2
Death Meas	Breast Cance	25.3	21.3
Death Meas	Colon Cancer	19.1	13.7
Death Meas	Coronary Hea	172	162

Demographics CSV



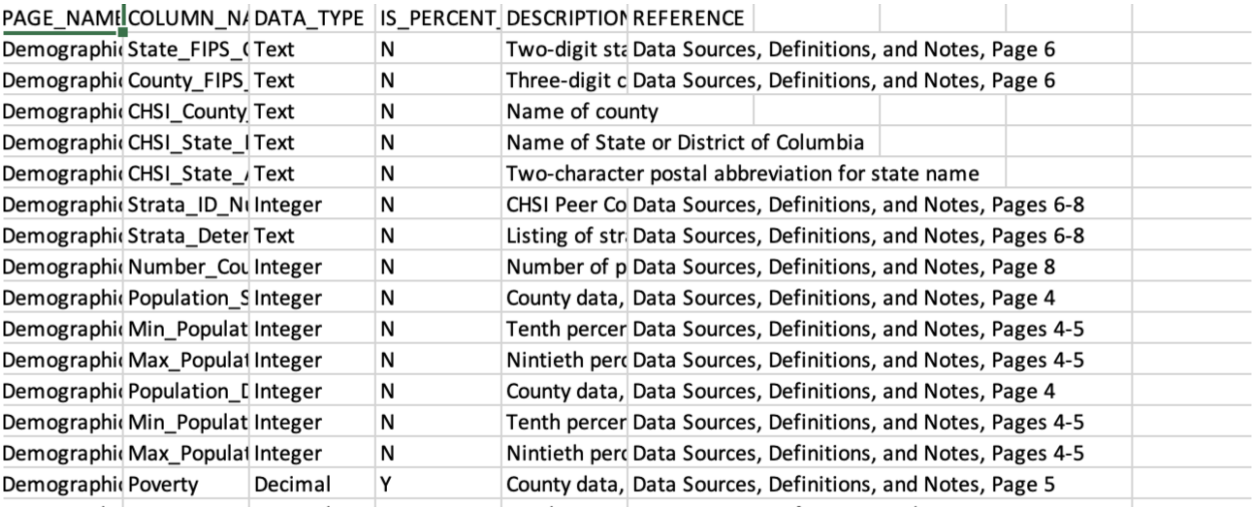
State_FIPS_	County_FIPS_	CHSI_County	CHSI_State_I	CHSI_State_/	Strata_ID_N	Strata_Deter	Number_Cou	Population_S	Min_Populat	Max_Populat	Population_[	Min_Populat	Max_Populat	Poverty	Min_Poverty	Max_Poverty	Age_19_Und	Min_Age_19
1	1	Autauga	Alabama	AL	29	frontier statu	37	48612	28447	55936	82	40	141	10.4	9.5	12.9	26.9	23.7
1	3	Baldwin	Alabama	AL	16	frontier statu	27	162586	118395	277035	102	39	457	10.2	9.7	12.9	23.5	21.3
1	5	Barbour	Alabama	AL	51	frontier statu	33	28414	27269	43226	32	14	41	22.1	18	24.6	24.3	23.5
1	7	Bibb	Alabama	AL	42	frontier statu	53	21516	8134	24778	35	9	66	16.8	12.5	16.4	24.6	24.4
1	9	Blount	Alabama	AL	28	frontier statu	39	55725	29009	53844	86	30	229	11.9	9.4	13.4	24.5	21.8
1	11	Bullock	Alabama	AL	75	frontier statu	37	11055	6228	19495	18	15	22	26.2	17	24.9	24.7	22.3
1	13	Butler	Alabama	AL	76	frontier statu	38	20766	9226	23786	27	24	42	20	16.7	23.3	25.6	24.8
1	15	Calhoun	Alabama	AL	6	frontier statu	53	112141	111380	231954	184	43	697	16.4	12.4	16.5	24.1	22.4
1	17	Chambers	Alabama	AL	50	frontier statu	27	35460	27028	48148	59	56	140	16.2	13.7	16.2	24.8	20.8
1	19	Cherokee	Alabama	AL	64	frontier statu	41	24522	9340	25391	44	35	46	15.2	12	15.7	21.9	19.6
1	21	Chilton	Alabama	AL	32	frontier statu	37	41744	29918	51327	60	25	310	14.9	12.7	17.8	25	20.8
1	23	Choctaw	Alabama	AL	66	frontier statu	37	14807	6709	17773	16	8	19	18.7	17.4	29.4	24.9	24.5
1	25	Clarke	Alabama	AL	51	frontier statu	33	27269	27269	43226	22	14	41	19.2	18	24.6	27.3	23.5
1	27	Clay	Alabama	AL	63	frontier statu	32	13964	9378	21479	23	24	32	14	12.1	16.5	22.4	21.1
1	29	Cleburne	Alabama	AL	41	frontier statu	47	14460	6602	24509	26	20	1362	14	12.5	17.5	23.3	21.8
1	31	Coffee	Alabama	AL	32	frontier statu	37	45567	29918	51327	67	25	310	13.7	12.7	17.8	24.5	20.8
1	33	Colbert	Alabama	AL	21	frontier statu	44	54660	53309	92614	92	40	201	14	9.8	13.5	23	20.6
1	35	Conecuh	Alabama	AL	75	frontier statu	37	13257	6228	19495	16	15	22	22	17	24.9	25.3	22.3
1	37	Coosa	Alabama	AL	41	frontier statu	47	11162	6602	24509	17	20	1362	13.4	12.5	17.5	23.2	21.8
1	39	Covington	Alabama	AL	35	frontier statu	27	37003	27228	49644	36	32	87	17.5	14.6	20.3	23.4	21.8
1	41	Crenshaw	Alabama	AL	71	frontier statu	33	13727	7147	20507	23	19	28	17.6	16.2	21.1	24.2	20

Defined Data Value CSV



Data_Value	Description
-9999	Indicate N.A. value from the source data for the Unemployed column on the VUNERABLEPOPSANDENVHEALTH page
-2222 or -22:	nda, no data available, see Data Notes document for details
-1111.1 or -1	nrf, no report, see Data Notes document for details
1	Represent 'No' in the indicator columns
2	Represent 'Yes' in the indicator columns
3	Represent 'Favorable to peers' in the indicator columns
4	Represent 'Unfavorable to peers' in the indicator columns
5	Represent ''Favorable to peers and favorable the U.S. Rate' in the indicator columns
6	Represent 'Favorable to peers and unfavorable the U.S. Rate' in the indicator columns
7	Represent 'Unfavorable to peers and favorable the U.S. Rate' in the indicator columns
8	Represent 'Unfavorable to peers and unfavorable the U.S. Rate' in the indicator columns
-9998.9	Indicate no objective for the Healthy People 2010 Target data

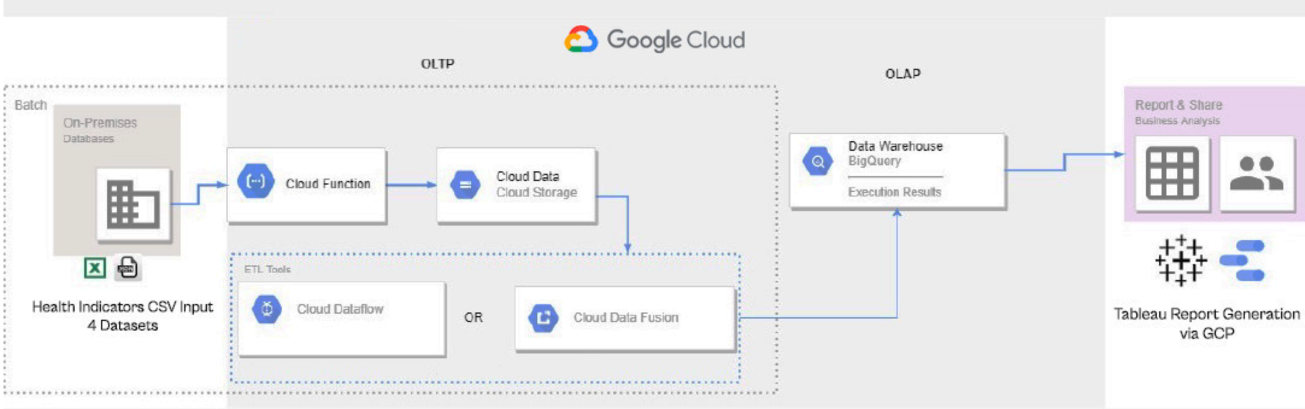
Data Element Description CSV



PAGE_NAME	COLUMN_N/	DATA_TYPE	IS_PERCENT_	DESCRIPTION	REFERENCE
Demographi	State_FIPS_(	Text	N	Two-digit sta	Data Sources, Definitions, and Notes, Page 6
Demographi	County_FIPS_	Text	N	Three-digit c	Data Sources, Definitions, and Notes, Page 6
Demographi	CHSI_County	Text	N	Name of county	
Demographi	CHSI_State_I	Text	N	Name of State or District of Columbia	
Demographi	CHSI_State_/	Text	N	Two-character postal abbreviation for state name	
Demographi	Strata_ID_Nı	Integer	N	CHSI Peer Co	Data Sources, Definitions, and Notes, Pages 6-8
Demographi	Strata_Deter	Text	N	Listing of str	Data Sources, Definitions, and Notes, Pages 6-8
Demographi	Number_Cou	Integer	N	Number of p	Data Sources, Definitions, and Notes, Page 8
Demographi	Population_S	Integer	N	County data,	Data Sources, Definitions, and Notes, Page 4
Demographi	Min_Populat	Integer	N	Tenth percer	Data Sources, Definitions, and Notes, Pages 4-5
Demographi	Max_Populat	Integer	N	Nintieth perc	Data Sources, Definitions, and Notes, Pages 4-5
Demographi	Population_[	Integer	N	County data,	Data Sources, Definitions, and Notes, Page 4
Demographi	Min_Populat	Integer	N	Tenth percer	Data Sources, Definitions, and Notes, Pages 4-5
Demographi	Max_Populat	Integer	N	Nintieth perc	Data Sources, Definitions, and Notes, Pages 4-5
Demographi	Poverty	Decimal	Y	County data,	Data Sources, Definitions, and Notes, Page 5

GOOGLE CLOUD DATABASE PIPELINE IMPLEMENTATION

Google Cloud Architecture Model

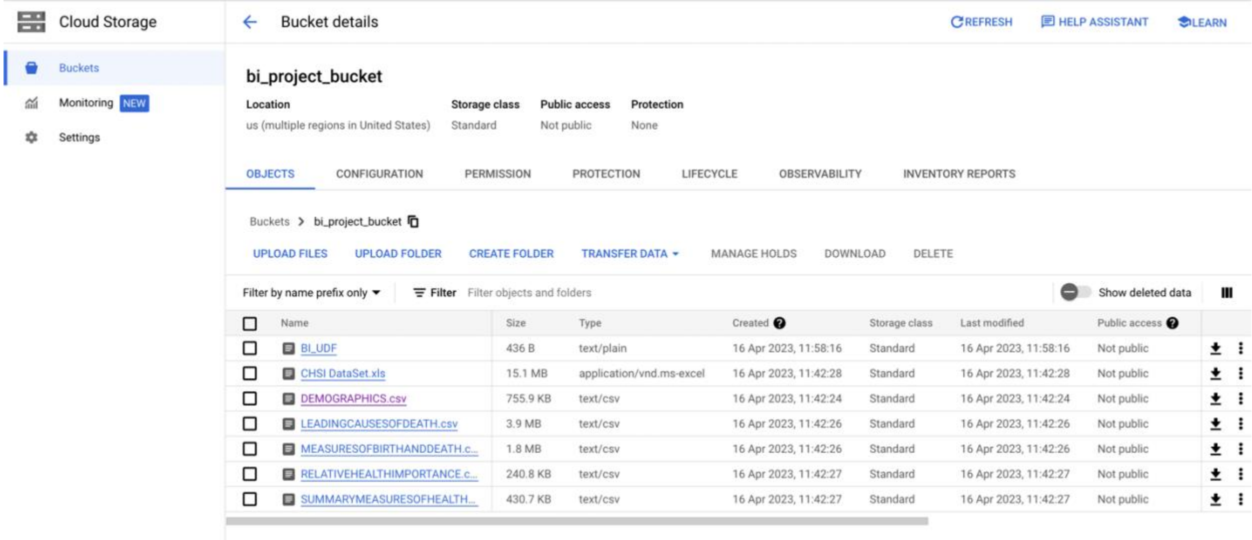


Health Indicator Data Warehouse over Google Cloud

Google Cloud Tools Used:
1.Google Cloud Storage
2.Google BigQuery
3.Google Cloud Data Fusion
4.SQL Workspace
5.IAM and Admin
6.Google Data Fusion – Wrangler
7.Google Cloud Shell

1. Google Cloud Storage:

Cloud Storage Bucket:



bi_project_bucket

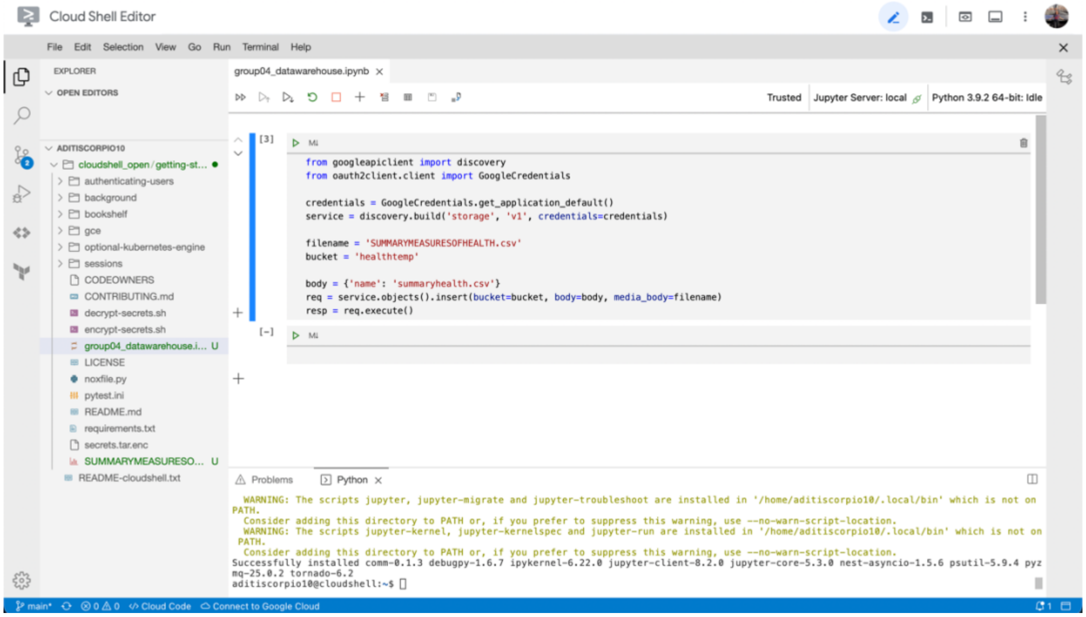
Location	Storage class	Public access	Protection
us (multiple regions in United States)	Standard	Not public	None

Name	Size	Type	Created	Storage class	Last modified	Public access
BI_UDF	436 B	text/plain	16 Apr 2023, 11:58:16	Standard	16 Apr 2023, 11:58:16	Not public
CHSI DataSet.xls	15.1 MB	application/vnd.ms-excel	16 Apr 2023, 11:42:28	Standard	16 Apr 2023, 11:42:28	Not public
DEMOGRAPHICS.csv	755.9 KB	text/csv	16 Apr 2023, 11:42:24	Standard	16 Apr 2023, 11:42:24	Not public
LEADINGCAUSESOFDEATH.csv	3.9 MB	text/csv	16 Apr 2023, 11:42:26	Standard	16 Apr 2023, 11:42:26	Not public
MEASURESOFBIRTHANDDEATH.c...	1.8 MB	text/csv	16 Apr 2023, 11:42:26	Standard	16 Apr 2023, 11:42:26	Not public
RELATIVEHEALTHIMPORTANCE.c...	240.8 KB	text/csv	16 Apr 2023, 11:42:27	Standard	16 Apr 2023, 11:42:27	Not public
SUMMARYMEASURESOFHEALTH...	430.7 KB	text/csv	16 Apr 2023, 11:42:27	Standard	16 Apr 2023, 11:42:27	Not public

We’ve inserted 5 datasets of our data warehouse “Health Indicators” into our BigQuery schema from their csv files using pipeline job in *Google Data Fusion Studio*

Using **Google Cloud Shell** to upload “Summary Health Indicator” dataset in Google Cloud Bucket “healthtemp” using Python programming language:

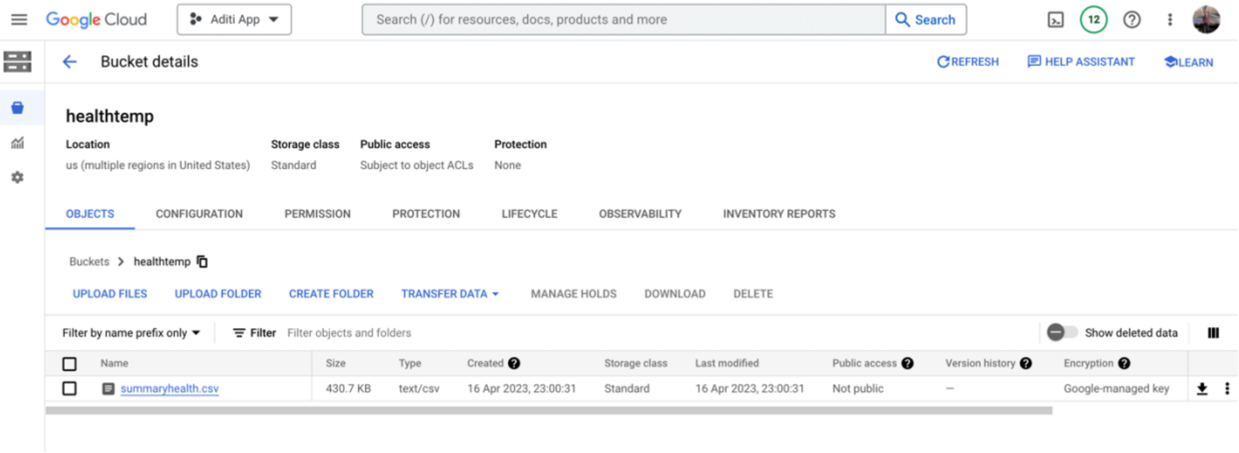
Google Cloud Shell .ipynb Python notebook-



```
[3]
from googleapiclient import discovery
from oauth2client.client import GoogleCredentials

credentials = GoogleCredentials.get_application_default()
service = discovery.build('storage', 'v1', credentials=credentials)

filename = 'SUMMARYMEASURESOFHEALTH.csv'
bucket = 'healthtemp'

body = {'name': 'summaryhealth.csv'}
req = service.objects().insert(bucket=bucket, body=body, media_body=filename)
resp = req.execute()
```

```
  WARNING: The scripts jupyter, jupyter-migrate and jupyter-troubleshoot are installed in '/home/aditiscorpio10/.local/bin' which is not on
PATH.
  Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-script-location.
  WARNING: The scripts jupyter-kernel, jupyter-kernelspec and jupyter-run are installed in '/home/aditiscorpio10/.local/bin' which is not on
 PATH.
  Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-script-location.
Successfully installed comm-0.1.3 debugpy-1.6.7 ipykernel-6.22.0 jupyter-client-8.2.0 jupyter-core-5.3.0 nest-asyncio-1.5.6 psutil-5.9.4 pyz
mq-25.0.2 tornado-6.2
aditiscorpio10@cloudshell:~$
```

“summaryhealth.csv” uploaded in “healthtemp” bucket-



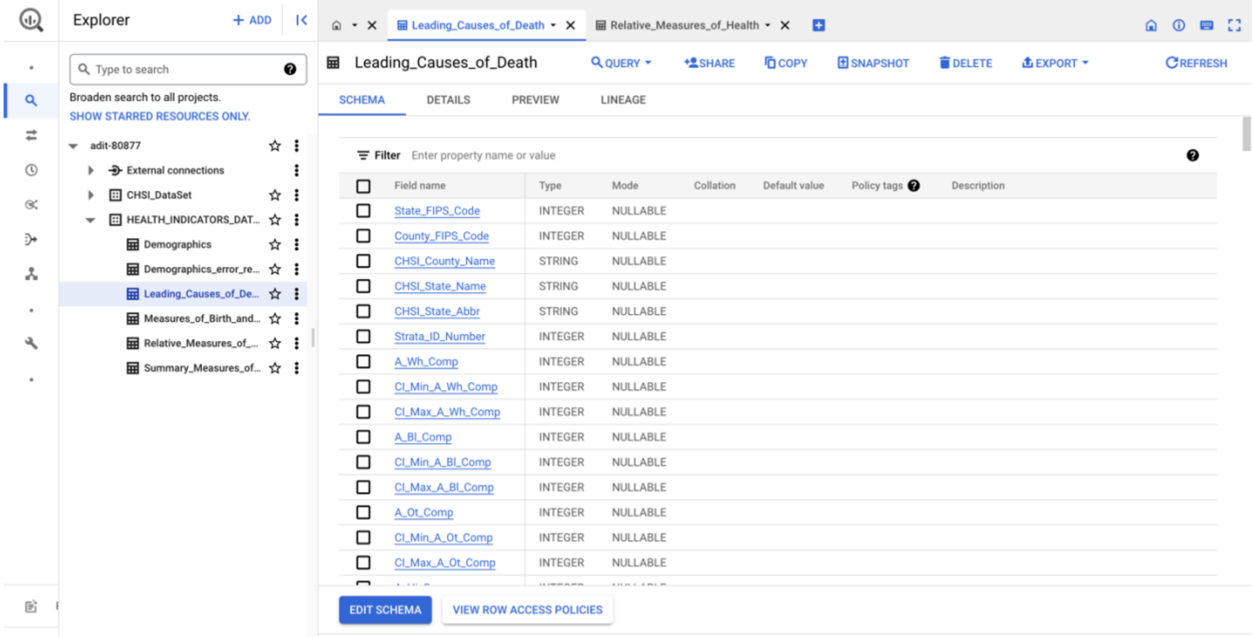
healthtemp

Location	Storage class	Public access	Protection
us (multiple regions in United States)	Standard	Subject to object ACLs	None

□	Name	Size	Type	Created	Storage class	Last modified	Public access	Version history	Encryption
□	summaryhealth.csv	430.7 KB	text/csv	16 Apr 2023, 23:00:31	Standard	16 Apr 2023, 23:00:31	Not public	–	Google-managed key

2. **Google BigQuery**

BigQuery Schema Tables:

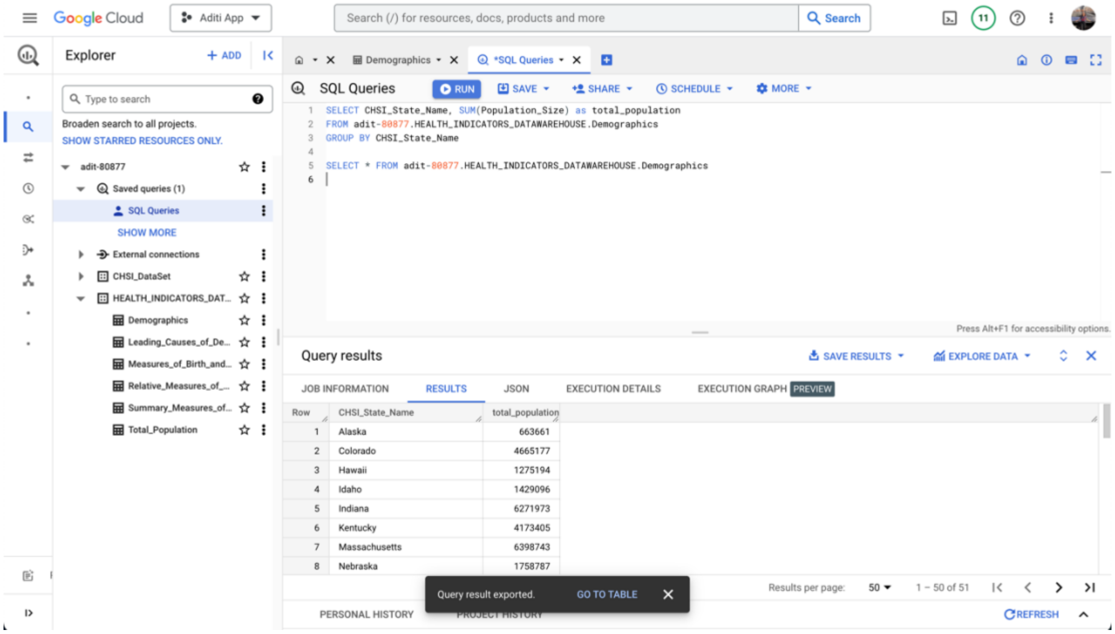


Leading_Causes_of_Death

	Field name	Type	Mode	Collation	Default value	Policy tags	Description
☐	State_FIPS_Code	INTEGER	NULLABLE				
☐	County_FIPS_Code	INTEGER	NULLABLE				
☐	CHSI_County_Name	STRING	NULLABLE				
☐	CHSI_State_Name	STRING	NULLABLE				
☐	CHSI_State_Abbr	STRING	NULLABLE				
☐	Strata_ID_Number	INTEGER	NULLABLE				
☐	A_Wh_Comp	INTEGER	NULLABLE				
☐	CI_Min_A_Wh_Comp	INTEGER	NULLABLE				
☐	CI_Max_A_Wh_Comp	INTEGER	NULLABLE				
☐	A_Bl_Comp	INTEGER	NULLABLE				
☐	CI_Min_A_Bl_Comp	INTEGER	NULLABLE				
☐	CI_Max_A_Bl_Comp	INTEGER	NULLABLE				
☐	A_Ot_Comp	INTEGER	NULLABLE				
☐	CI_Min_A_Ot_Comp	INTEGER	NULLABLE				
☐	CI_Max_A_Ot_Comp	INTEGER	NULLABLE				

Loaded datasets into data warehouse “Health Indicators” with 5 data tables- Demographics, Leading Causes of Death, Measures of Birth and Death, Relative Measures of health and Summary Measures of Health

Using *SQL Workspace* in BigQuery to GROUPBY and concatenate populations of different states into one column → exporting it in BigQuery as Table in the “Health Indicator” data warehouse

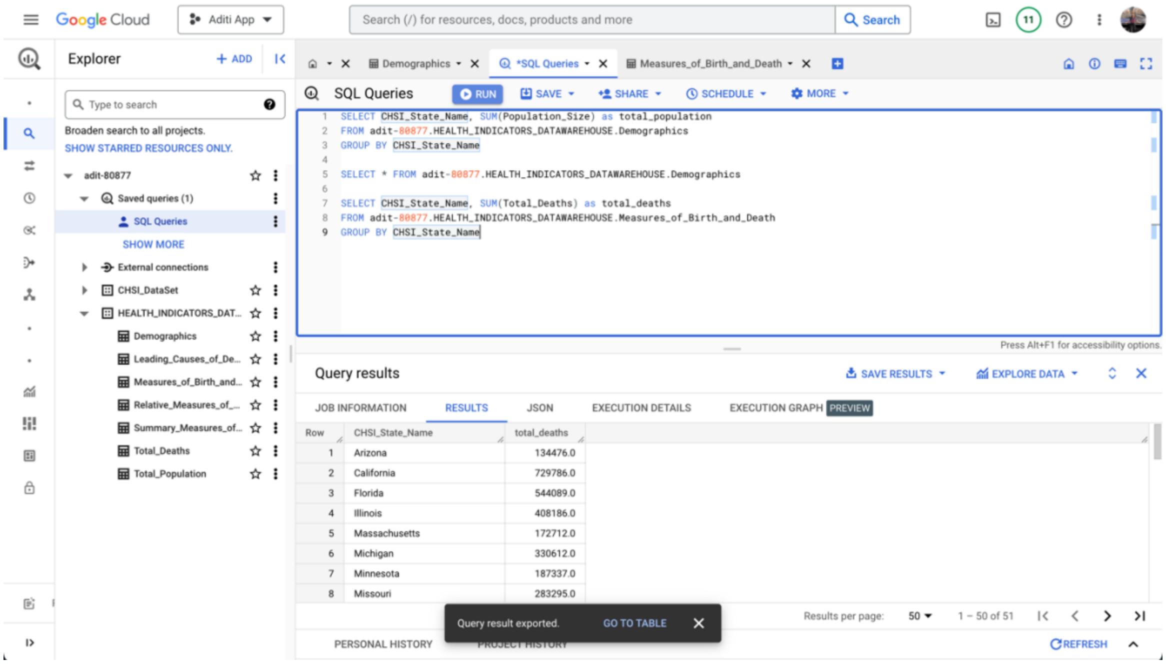


```
SELECT CHSI_State_Name, SUM(Population_Size) as total_population
FROM adit-80877.HEALTH_INDICATORS_DATAWAREHOUSE.Demographics
GROUP BY CHSI_State_Name

SELECT * FROM adit-80877.HEALTH_INDICATORS_DATAWAREHOUSE.Demographics
```

Row	CHSI_State_Name	total_population
1	Alaska	663661
2	Colorado	4665177
3	Hawaii	1275194
4	Idaho	1429096
5	Indiana	6271973
6	Kentucky	4173405
7	Massachusetts	6398743
8	Nebraska	1758787

Query result exported. GO TO TABLE

On *SQL Workspace* in BigQuery, performing GROUPBY on Total_Deaths in Measures_of_Births_and_Death dataset → exporting it in BigQuery as Total_Deaths table in the "Health Indicator" data warehouse



```
SELECT CHSI_State_Name, SUM(Population_Size) as total_population
FROM adit-80877.HEALTH_INDICATORS_DATAWAREHOUSE.Demographics
GROUP BY CHSI_State_Name

SELECT * FROM adit-80877.HEALTH_INDICATORS_DATAWAREHOUSE.Demographics

SELECT CHSI_State_Name, SUM(Total_Deaths) as total_deaths
FROM adit-80877.HEALTH_INDICATORS_DATAWAREHOUSE.Measures_of_Birth_and_Death
GROUP BY CHSI_State_Name
```

Query results

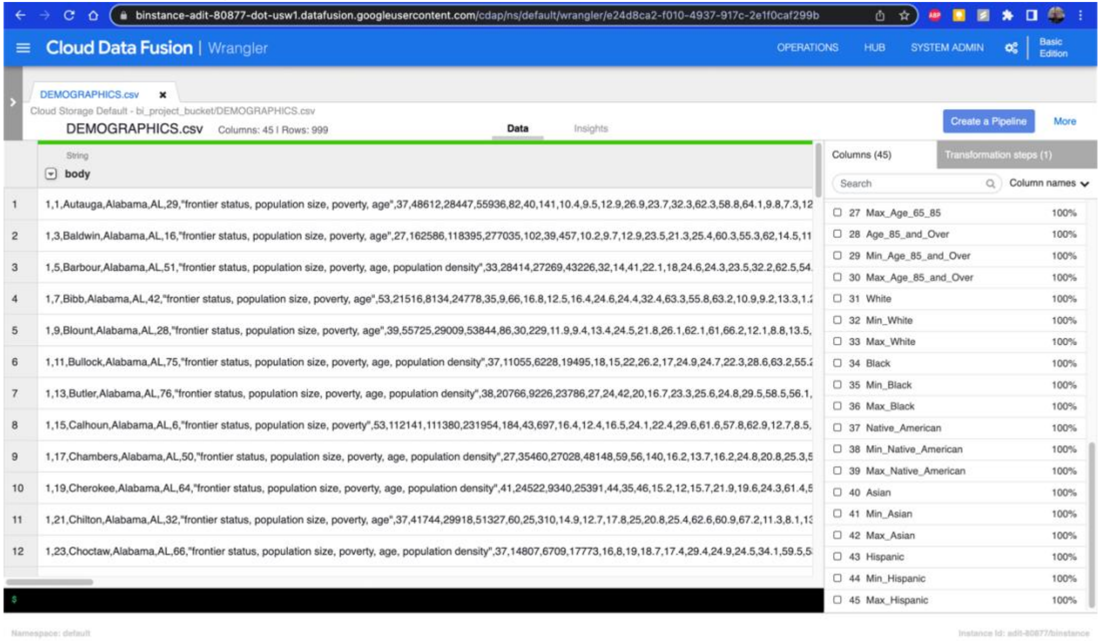
Row	CHSI_State_Name	total_deaths
1	Arizona	134476.0
2	California	729786.0
3	Florida	544089.0
4	Illinois	408186.0
5	Massachusetts	172712.0
6	Michigan	330612.0
7	Minnesota	187337.0
8	Missouri	283295.0

Query result exported.

Results per page: 50 1 – 50 of 51

3. Google Cloud Data Fusion- Wrangler

Text file loading into Data Fusion Studio



binstance-adit-80877-dot-usw1.datafusion.googleusercontent.com/cdap/ns/default/wrangler/e24d8ca2-f010-4937-917c-2e1f0caf299b
Cloud Data Fusion | Wrangler
OPERATIONS
HUB
SYSTEM ADMIN
Basic Edition
DEMOGRAPHICS.csv
Data
Insights
Create a Pipeline
More
Transformation steps (1)
Search
Column names
Namespace: default
Instance Id: adit-80877/binstance

Cloud Storage Default - bi_project_bucket/DEMOGRAPHICS.csv

DEMOGRAPHICS.csv Columns: 45 | Rows: 999

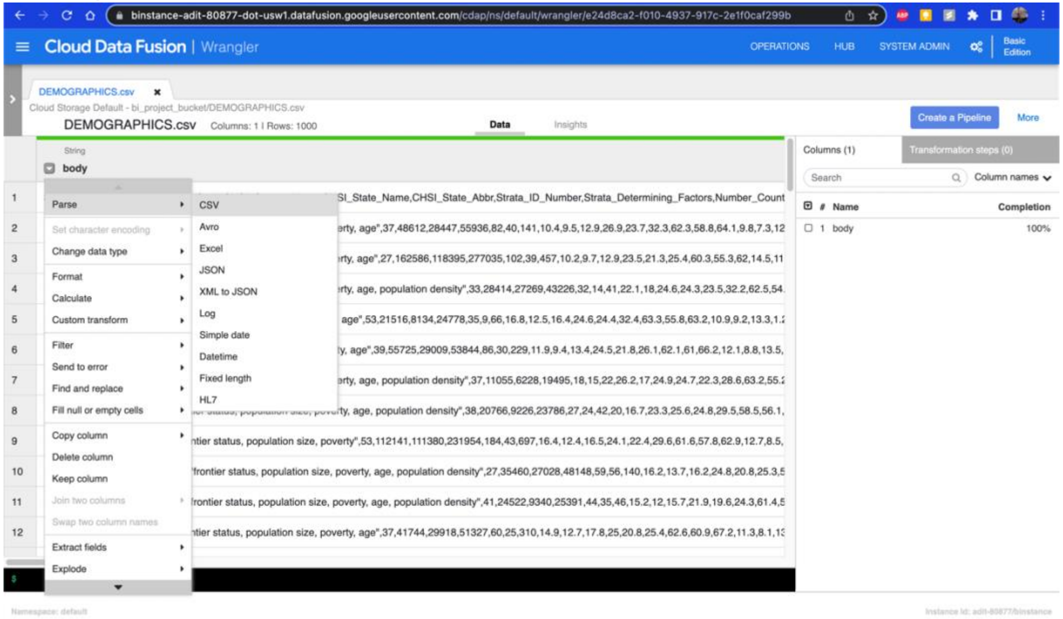
	String body
1	1,1,Autauga,Alabama,AL,29,"frontier status, population size, poverty, age",37,48612,28447,55936,82,40,141,10.4,9.5,12.9,26.9,23.7,32.3,62.3,58.8,64.1,9.8,7.3,12
2	1,3,Baldwin,Alabama,AL,16,"frontier status, population size, poverty, age",27,162586,118395,277035,102,39,457,10.2,9.7,12.9,23.5,21.3,25.4,60.3,55.3,62,14.5,11
3	1,5,Barbour,Alabama,AL,51,"frontier status, population size, poverty, age, population density",33,28414,27269,43226,32,14,41,22.1,18,24.6,24.3,23.5,32.2,62.5,54.
4	1,7,Bibb,Alabama,AL,42,"frontier status, population size, poverty, age",53,21516,8134,24778,35,9,66,16.8,12.5,16.4,24.6,24.4,32.4,63.3,55.8,63.2,10.9,9.2,13.3,1.
5	1,9,Blount,Alabama,AL,28,"frontier status, population size, poverty, age",39,55725,29009,53844,86,30,229,11.9,9.4,13.4,24.5,21.8,26.1,62.1,61,66.2,12.1,8.8,13.5,
6	1,11,Bullock,Alabama,AL,75,"frontier status, population size, poverty, age, population density",37,11055,6228,19495,18,15,22,26.2,17,24.9,24.7,22.3,28.6,63.2,55.
7	1,13,Butler,Alabama,AL,76,"frontier status, population size, poverty, age, population density",38,20766,9226,23786,27,24,42,20,16.7,23.3,25.6,24.8,29.5,58.5,56.1,
8	1,15,Calhoun,Alabama,AL,6,"frontier status, population size, poverty",53,112141,111380,231954,184,43,697,16.4,12.4,16.5,24.1,22.4,29.6,61.6,57.8,62.9,12.7,8.5,
9	1,17,Chambers,Alabama,AL,50,"frontier status, population size, poverty, age, population density",27,35460,27028,48148,59,56,140,16.2,13.7,16.2,24.8,20.8,25.3,5
10	1,19,Cherokee,Alabama,AL,64,"frontier status, population size, poverty, age, population density",41,24522,9340,25391,44,35,46,15.2,12,15.7,21.9,19.6,24.3,61.4,5
11	1,21,Chilton,Alabama,AL,32,"frontier status, population size, poverty, age",37,41744,29918,51327,60,25,310,14.9,12.7,17.8,25,20.8,25.4,62.6,60.9,67.2,11.3,8.1,13
12	1,23,Choctaw,Alabama,AL,66,"frontier status, population size, poverty, age, population density",37,14807,6709,17773,16.8,19,18.7,17.4,29.4,24.9,24.5,34.1,59.5,5

$

Columns (45)

□ 27 Max_Age_65_85	100%
□ 28 Age_85_and_Over	100%
□ 29 Min_Age_85_and_Over	100%
□ 30 Max_Age_85_and_Over	100%
□ 31 White	100%
□ 32 Min_White	100%
□ 33 Max_White	100%
□ 34 Black	100%
□ 35 Min_Black	100%
□ 36 Max_Black	100%
□ 37 Native_American	100%
□ 38 Min_Native_American	100%
□ 39 Max_Native_American	100%
□ 40 Asian	100%
□ 41 Min_Asian	100%
□ 42 Max_Asian	100%
□ 43 Hispanic	100%
□ 44 Min_Hispanic	100%
□ 45 Max_Hispanic	100%

#Tranformation Step 1:
Parsing it as CSV



binstance-adit-80877-dot-usw1.datafusion.googleusercontent.com/cdap/ns/default/wrangler/e24d8ca2-f010-4937-917c-2e1f0caf299b

Cloud Storage Default - bi_project_bucket/DEMOGRAPHICS.csv
DEMOGRAPHICS.csv Columns: 1 | Rows: 1000

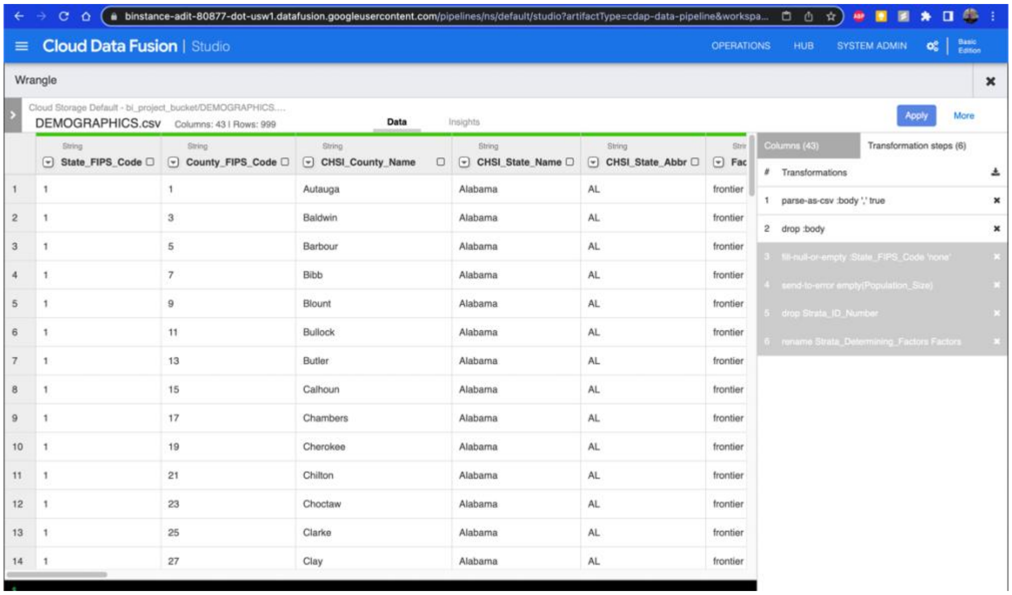
	String body
1	SI_State_Name,CHSI_State_Abbr,Strata_ID_Number,Strata_Determining_Factors,Number_Count
2	erty, age",37,48612,28447,55936,82,40,141,10.4,9.5,12.9,26.9,23.7,32.3,62.3,58.8,64.1,9.8,7.3,12
3	erty, age",27,162586,118395,277035,102,39,457,10.2,9.7,12.9,23.5,21.3,25.4,60.3,55.3,62,14.5,11
4	erty, age, population density",33,28414,27269,43226,32,14,41,22.1,18,24.6,24.3,23.5,32.2,62.5,54.
5	age",53,21516,8134,24778,35,9,66,16.8,12.5,16.4,24.6,24.4,32.4,63.3,55.8,63.2,10.9,9.2,13.3,1.2
6	y, age",39,55725,29009,53844,86,30,229,11.9,9.4,13.4,24.5,21.8,26.1,62.1,61,66.2,12.1,8.8,13.5,
7	erty, age, population density",37,11055,6228,19495,18,15,22,26.2,17,24.9,24.7,22.3,28.6,63.2,55.2
8	ty, age, population density",38,20766,9226,23786,27,24,42,20,16.7,23.3,25.6,24.8,29.5,58.5,56.1,
9	ntier status, population size, poverty",53,112141,111380,231954,184,43,697,16.4,12.4,16.5,24.1,22.4,29.6,61.6,57.8,62.9,12.7,8.5,
10	'frontier status, population size, poverty, age, population density",27,35460,27028,48148,59,56,140,16.2,13.7,16.2,24.8,20.8,25.3,5
11	frontier status, population size, poverty, age, population density",41,24522,9340,25391,44,35,46,15.2,12,15.7,21.9,19.6,24.3,61.4,5
12	ntier status, population size, poverty, age",37,41744,29918,51327,60,25,310,14.9,12.7,17.8,25,20.8,25.4,62.6,60.9,67.2,11.3,8.1,13

#	Name	Completion
1	body	100%

#Tranformation Steps 2-6:

Adding other transformation steps of data cleaning and renaming in Google Cloud Console command line as Recipe for Wrangling as-

```
fill-null-or-empty :State_FIPS_Code 'none'
send-to-error empty(Population_Size)
drop Strata_ID_Number
rename Strata_Determining_Factors Factors
```



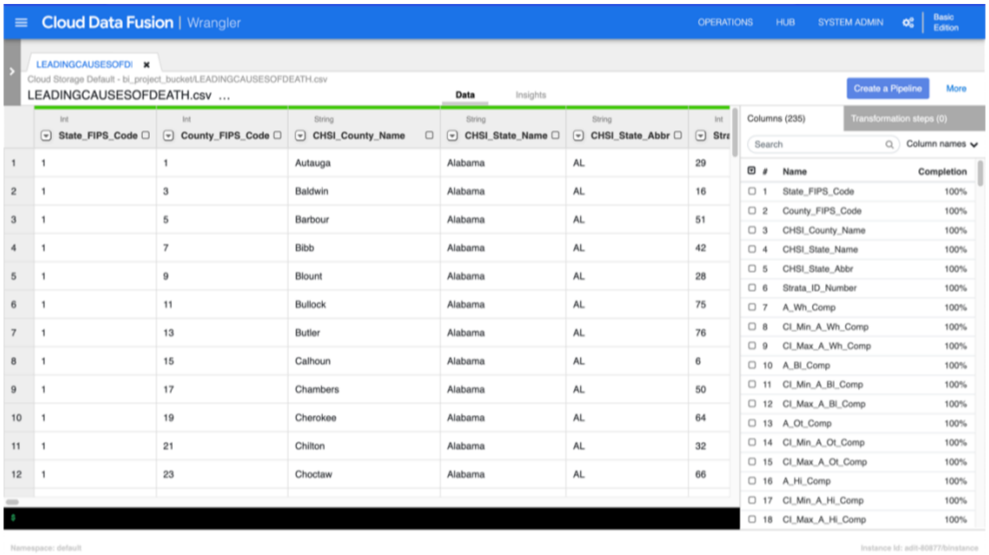
binstance-adit-80877-dot-usw1.datafusion.googleusercontent.com/pipelines/ns/default/studio?artifactType=cdap-data-pipeline&workspa...

Cloud Storage Default - bi_project_bucket/DEMOGRAPHICS....
DEMOGRAPHICS.csv Columns: 43 | Rows: 999

	String State_FIPS_Code	String County_FIPS_Code	String CHSI_County_Name	String CHSI_State_Name	String CHSI_State_Abbr	String Fac
1	1	1	Autauga	Alabama	AL	frontier
2	1	3	Baldwin	Alabama	AL	frontier
3	1	5	Barbour	Alabama	AL	frontier
4	1	7	Bibb	Alabama	AL	frontier
5	1	9	Blount	Alabama	AL	frontier
6	1	11	Bullock	Alabama	AL	frontier
7	1	13	Butler	Alabama	AL	frontier
8	1	15	Calhoun	Alabama	AL	frontier
9	1	17	Chambers	Alabama	AL	frontier
10	1	19	Cherokee	Alabama	AL	frontier
11	1	21	Chilton	Alabama	AL	frontier
12	1	23	Choctaw	Alabama	AL	frontier
13	1	25	Clarke	Alabama	AL	frontier
14	1	27	Clay	Alabama	AL	frontier

#	Transformations
1	parse-as-csv :body ',' true
2	drop :body
3	fill-null-or-empty :State_FIPS_Code 'none'
4	send-to-error empty(Population_Size)
5	drop Strata_ID_Number
6	rename Strata_Determining_Factors Factors

Similarly transforming datasets 2-5:



Cloud Storage Default - bi_project_bucket/LEADINGCAUSESOFDEATH.csv

LEADINGCAUSESOFDEATH.csv ...

	Int State_FIPS_Code	Int County_FIPS_Code	String CHSI_County_Name	String CHSI_State_Name	String CHSI_State_Abbr	Int Stra
1	1	1	Autauga	Alabama	AL	29
2	1	3	Baldwin	Alabama	AL	16
3	1	5	Barbour	Alabama	AL	51
4	1	7	Bibb	Alabama	AL	42
5	1	9	Blount	Alabama	AL	28
6	1	11	Bullock	Alabama	AL	75
7	1	13	Butler	Alabama	AL	76
8	1	15	Calhoun	Alabama	AL	6
9	1	17	Chambers	Alabama	AL	50
10	1	19	Cherokee	Alabama	AL	64
11	1	21	Chilton	Alabama	AL	32
12	1	23	Choctaw	Alabama	AL	66

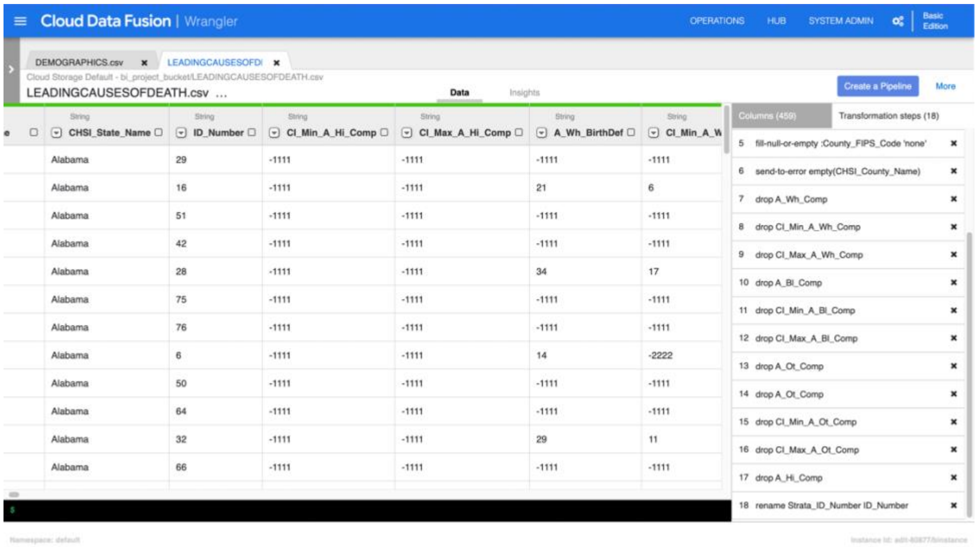
Columns (235)

#	Name	Completion
1	State_FIPS_Code	100%
2	County_FIPS_Code	100%
3	CHSI_County_Name	100%
4	CHSI_State_Name	100%
5	CHSI_State_Abbr	100%
6	Strata_ID_Number	100%
7	A_Wh_Comp	100%
8	CI_Min_A_Wh_Comp	100%
9	CI_Max_A_Wh_Comp	100%
10	A_Bl_Comp	100%
11	CI_Min_A_Bl_Comp	100%
12	CI_Max_A_Bl_Comp	100%
13	A_Ot_Comp	100%
14	CI_Min_A_Ot_Comp	100%
15	CI_Max_A_Ot_Comp	100%
16	A_Hi_Comp	100%
17	CI_Min_A_Hi_Comp	100%
18	CI_Max_A_Hi_Comp	100%

#Transformation Steps

Wrangling Recipe for 18 transformation steps-

```
fill-null-or-empty :County_FIPS_Code 'none'
send-to-error empty(CHSI_County_Name)
drop A_Wh_Comp
drop CI_Min_A_Wh_Comp
drop CI_Max_A_Wh_Comp
drop A_Bl_Comp
drop CI_Min_A_Bl_Comp
drop CI_Max_A_Bl_Comp
drop A_Ot_Comp
drop A_Ot_Comp
drop CI_Min_A_Ot_Comp
drop CI_Max_A_Ot_Comp
drop A_Hi_Comp
rename Strata_ID_Number ID_Number
```



Cloud Data Fusion | Wrangler

DEMOGRAPHICS.csv ✖ LEADINGCAUSESOFDI ✖

Cloud Storage Default - bi_project_bucket/LEADINGCAUSESOFDEATH.csv

LEADINGCAUSESOFDEATH.csv …

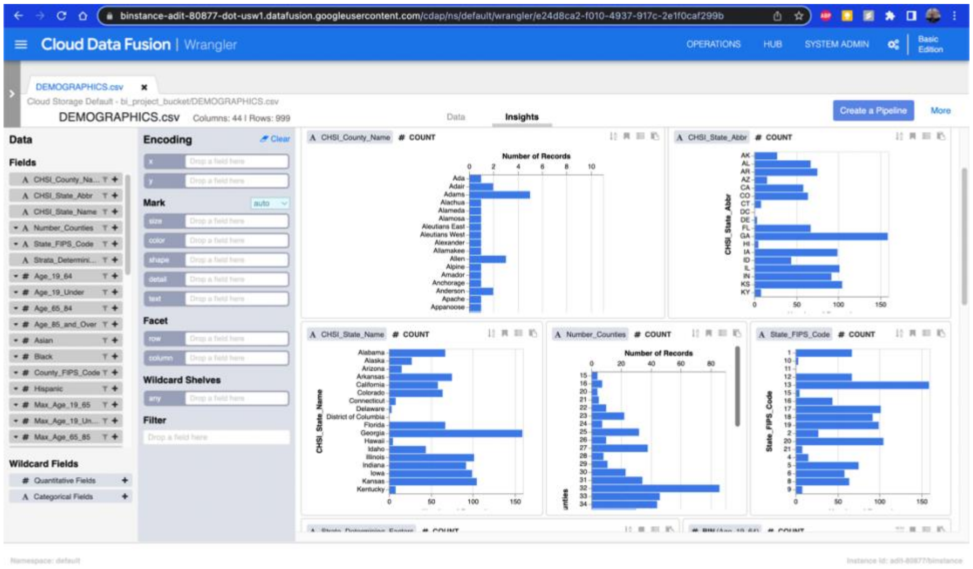
CHSI_State_Name	ID_Number	CI_Min_A_Hi_Comp	CI_Max_A_Hi_Comp	A_Wh_BirthDef	CI_Min_A_W
String	String	String	String	String	String
Alabama	29	-1111	-1111	-1111	-1111
Alabama	16	-1111	-1111	21	6
Alabama	51	-1111	-1111	-1111	-1111
Alabama	42	-1111	-1111	-1111	-1111
Alabama	28	-1111	-1111	34	17
Alabama	75	-1111	-1111	-1111	-1111
Alabama	76	-1111	-1111	-1111	-1111
Alabama	6	-1111	-1111	14	-2222
Alabama	50	-1111	-1111	-1111	-1111
Alabama	64	-1111	-1111	-1111	-1111
Alabama	32	-1111	-1111	29	11
Alabama	66	-1111	-1111	-1111	-1111

Columns (459) Transformation steps (18)

5 fill-null-or-empty :County_FIPS_Code 'none'
6 send-to-error empty(CHSI_County_Name)
7 drop A_Wh_Comp
8 drop CI_Min_A_Wh_Comp
9 drop CI_Max_A_Wh_Comp
10 drop A_Bl_Comp
11 drop CI_Min_A_Bl_Comp
12 drop CI_Max_A_Bl_Comp
13 drop A_Ot_Comp
14 drop A_Ot_Comp
15 drop CI_Min_A_Ot_Comp
16 drop CI_Max_A_Ot_Comp
17 drop A_Hi_Comp
18 rename Strata_ID_Number ID_Number

Analytics Dashboard for Datasets-

Demographics:



binstance-adit-80877-dot-usw1.datafusion.googleusercontent.com/cdap/ns/default/wrangler/e24d8ca2-f010-4937-917c-2e1f0caf299b

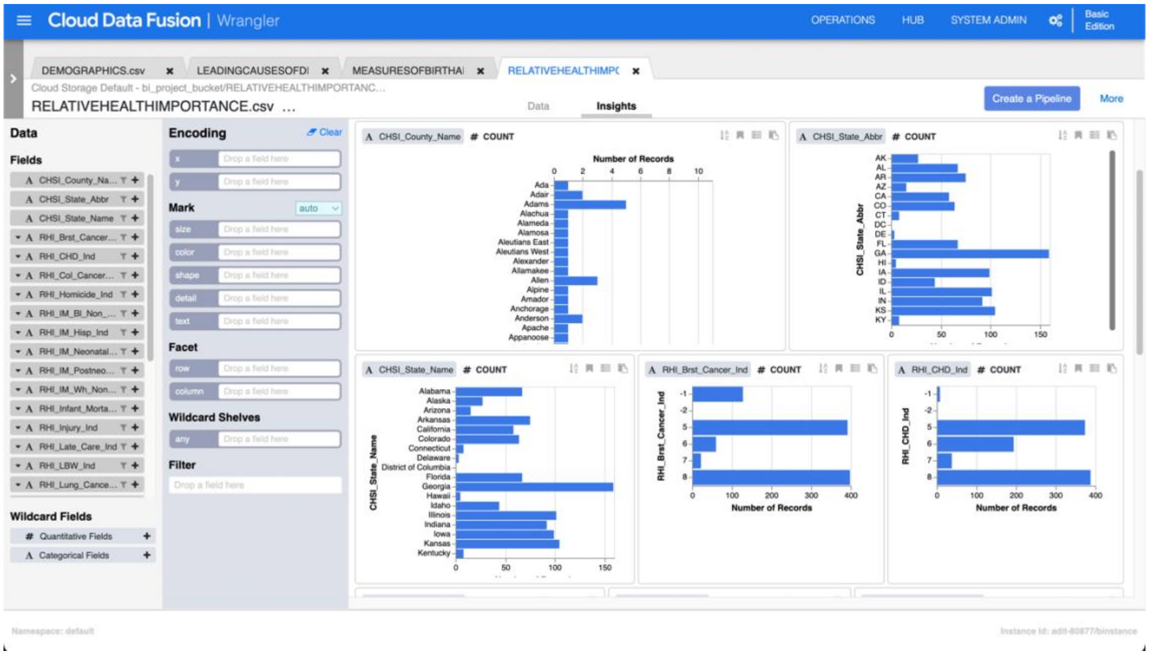
Cloud Data Fusion | Wrangler

DEMOGRAPHICS.csv ✖

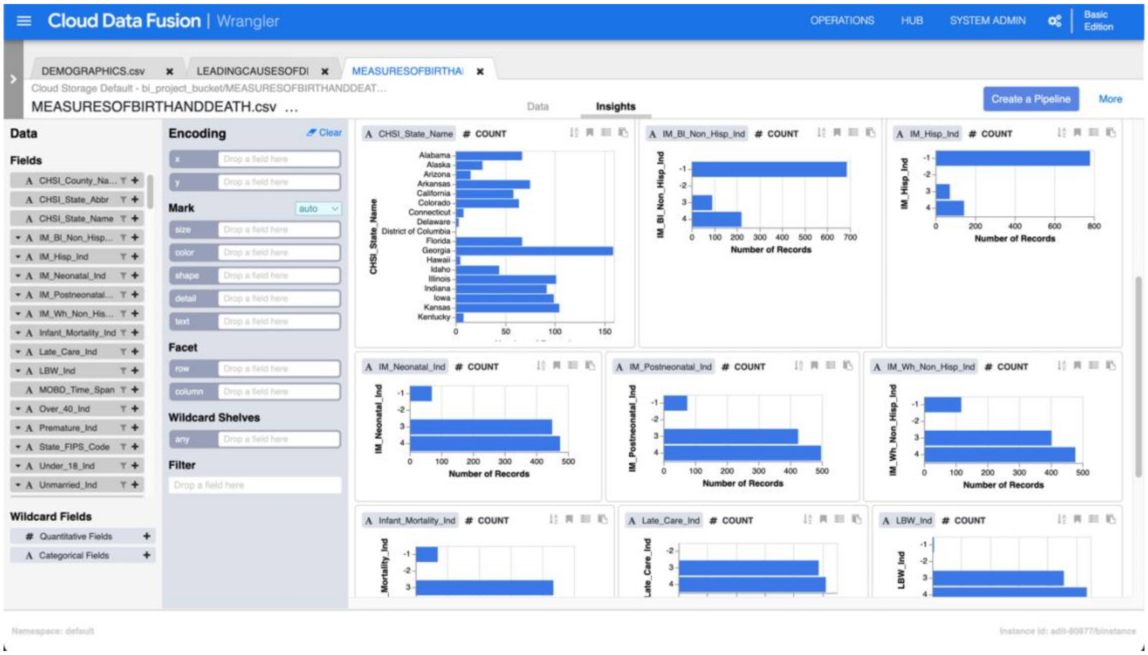
Cloud Storage Default - bi_project_bucket/DEMOGRAPHICS.csv

DEMOGRAPHICS.csv Columns: 44 | Rows: 999

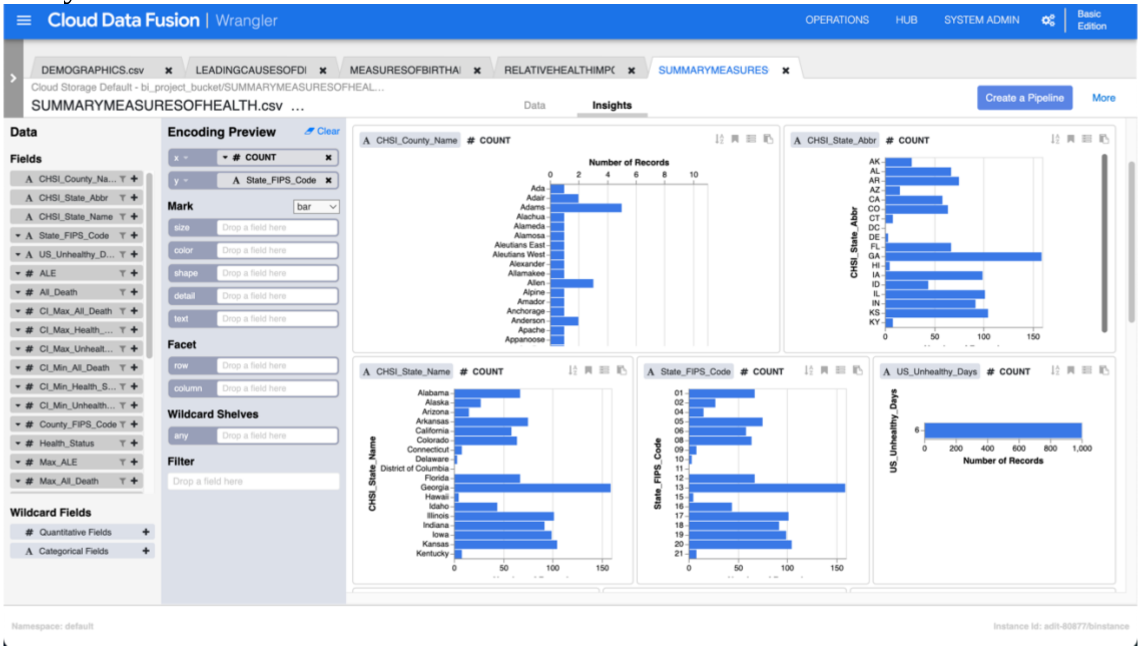
Relative Measures of Health-



Measures of Birth and Death-



Summary Measures of Health-



Cloud Data Fusion | Wrangler

Cloud Storage Default - bi_project_bucket/SUMMARYMEASURESOFHEAL...

SUMMARYMEASURESOFHEALTH.csv ...

Encoding Preview

x: # COUNT

y: A State_FIPS_Code

Mark: bar

Namespace: default

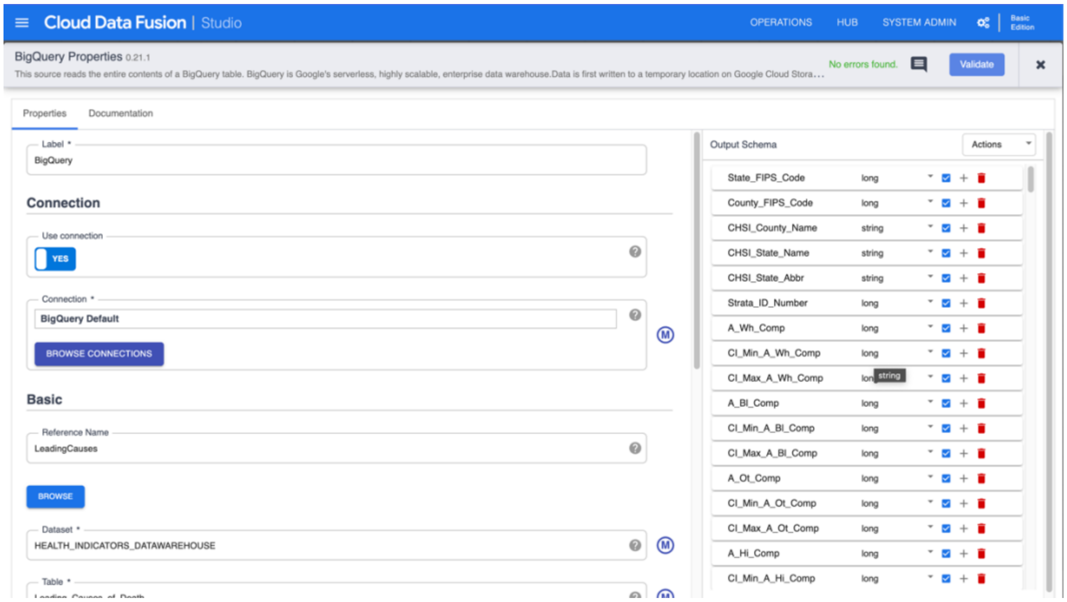
Instance Id: adit-80877/binstance

Data Fusion Studio: Job Pipeline Creation

Google Cloud File → Data Wrangling → Storing in BigQuery Data Warehouse Schema

Extract → Transform → Load

Setting up BigQuery with our bucket "bi_bucket" as location-



Cloud Data Fusion | Studio

BigQuery Properties 0.21.1

This source reads the entire contents of a BigQuery table. BigQuery is Google's serverless, highly scalable, enterprise data warehouse.Data is first written to a temporary location on Google Cloud Stora...

No errors found.

Label *: BigQuery

Connection

Use connection: YES

Connection *: BigQuery Default

Basic

Reference Name: LeadingCauses

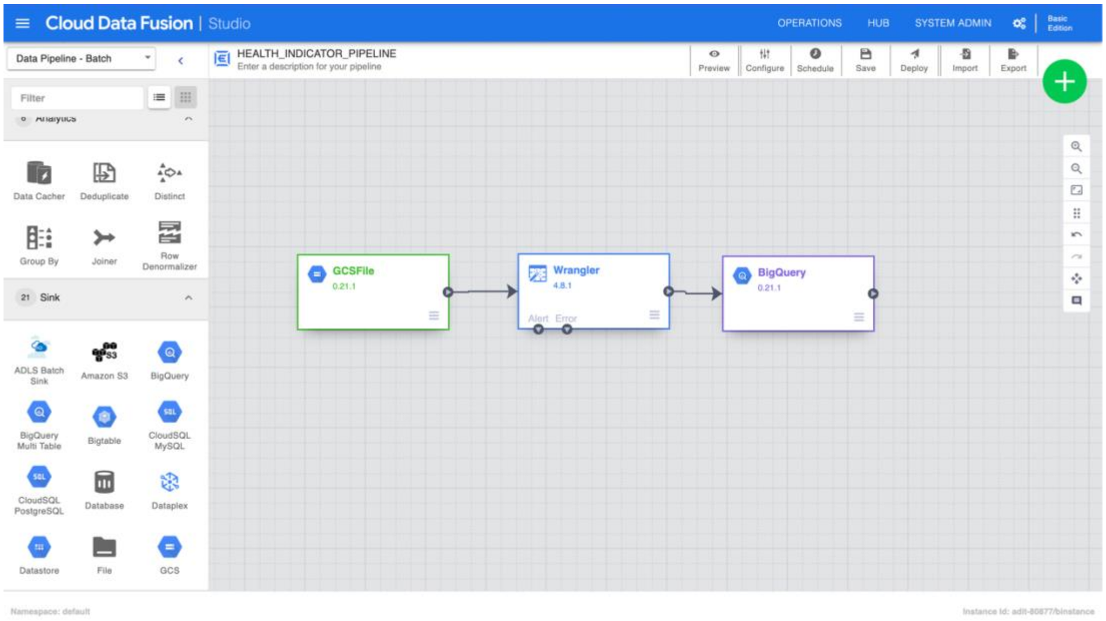
Dataset *: HEALTH_INDICATORS_DATAWAREHOUSE

Table *: Leading_Causes_of_Death

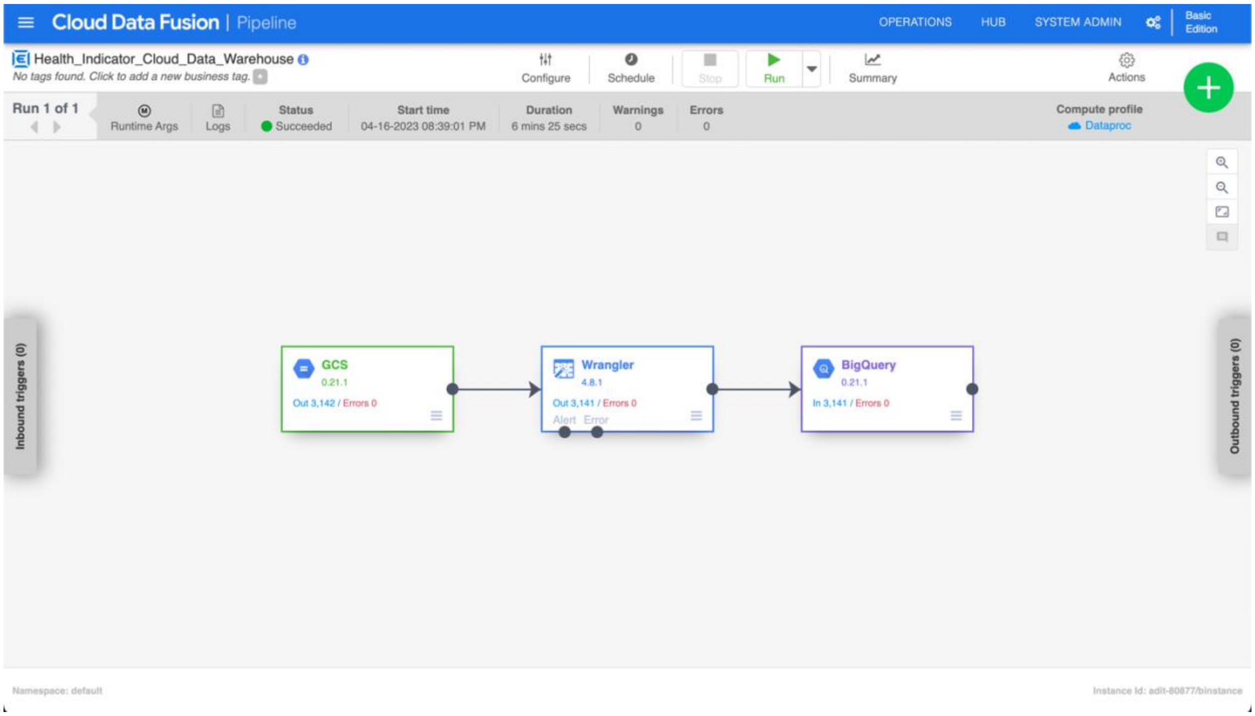
Output Schema

State_FIPS_Code	long
County_FIPS_Code	long
CHSI_County_Name	string
CHSI_State_Name	string
CHSI_State_Abbr	string
Strata_ID_Number	long
A_Wh_Comp	long
CI_Min_A_Wh_Comp	long
CI_Max_A_Wh_Comp	lon string
A_Bl_Comp	long
CI_Min_A_Bl_Comp	long
CI_Max_A_Bl_Comp	long
A_Ot_Comp	long
CI_Min_A_Ot_Comp	long
CI_Max_A_Ot_Comp	long
A_Hi_Comp	long
CI_Min_A_Hi_Comp	long

Database Pipeline Connection as “Health Indicator Pipeline”-

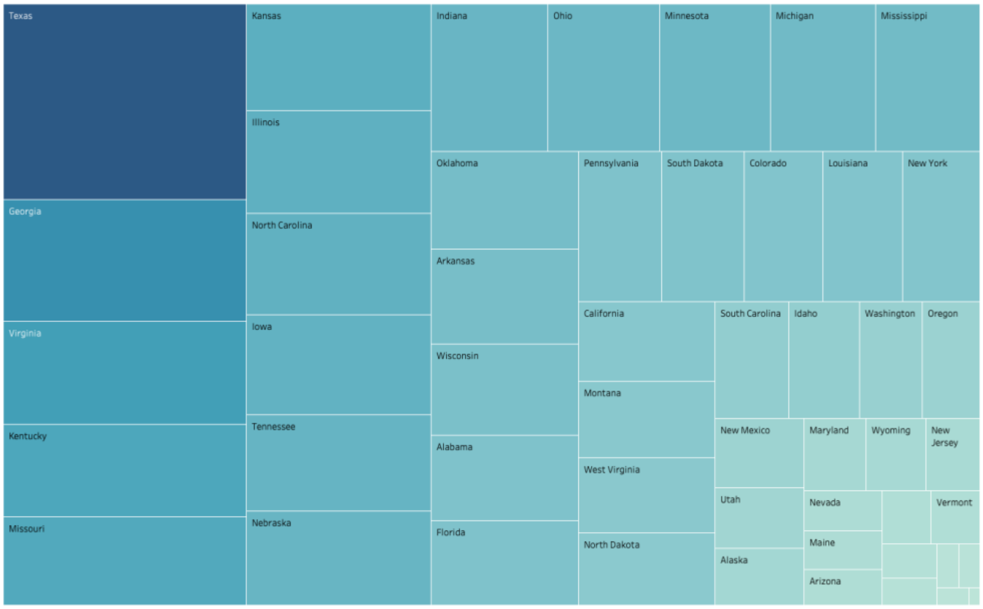


Successful database pipeline job run saving 3,141 cleaned data entries into “bi_bucket”/*BigQuery DataWarehouse Schema*-


Cloud Data Fusion | Pipeline
OPERATIONS
HUB
SYSTEM ADMIN
Basic Edition
Health_Indicator_Cloud_Data_Warehouse
No tags found. Click to add a new business tag.
Configure
Schedule
Stop
Run
Summary
Actions
Run 1 of 1
Runtime Args
Logs

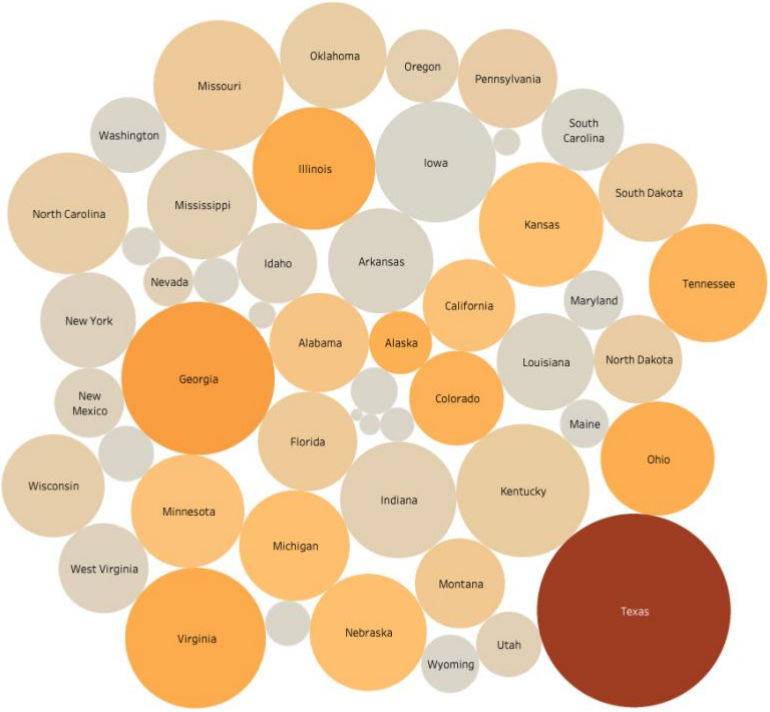
Status: Succeeded
Start time: 04-16-2023 08:39:01 PM
Duration: 6 mins 25 secs
Warnings: 0
Errors: 0
Compute profile: Dataproc

ANALYSIS – KPI DASHBOARDS – Tableau Connection to Google Cloud



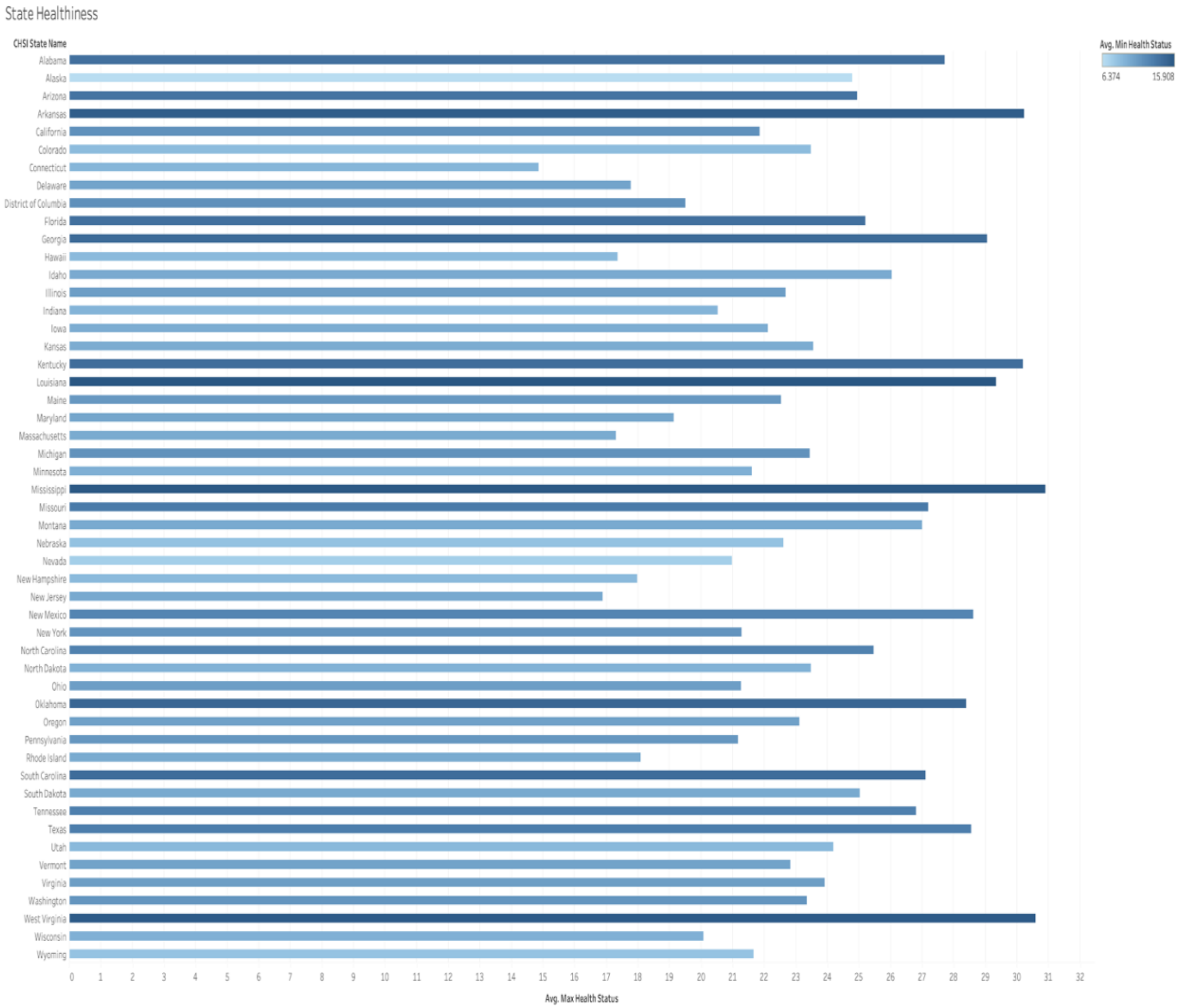
Heart Diseases Tree map

Total deaths causes by Heart Diseases are highest in TEXAS



Overall total deaths in the US

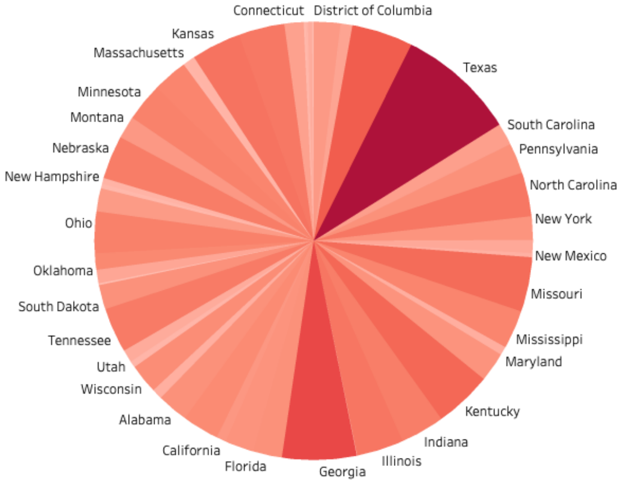
Highest is in Texas → Georgia → Virginia → Illinois → Kansas → Tennessee → Nebraska



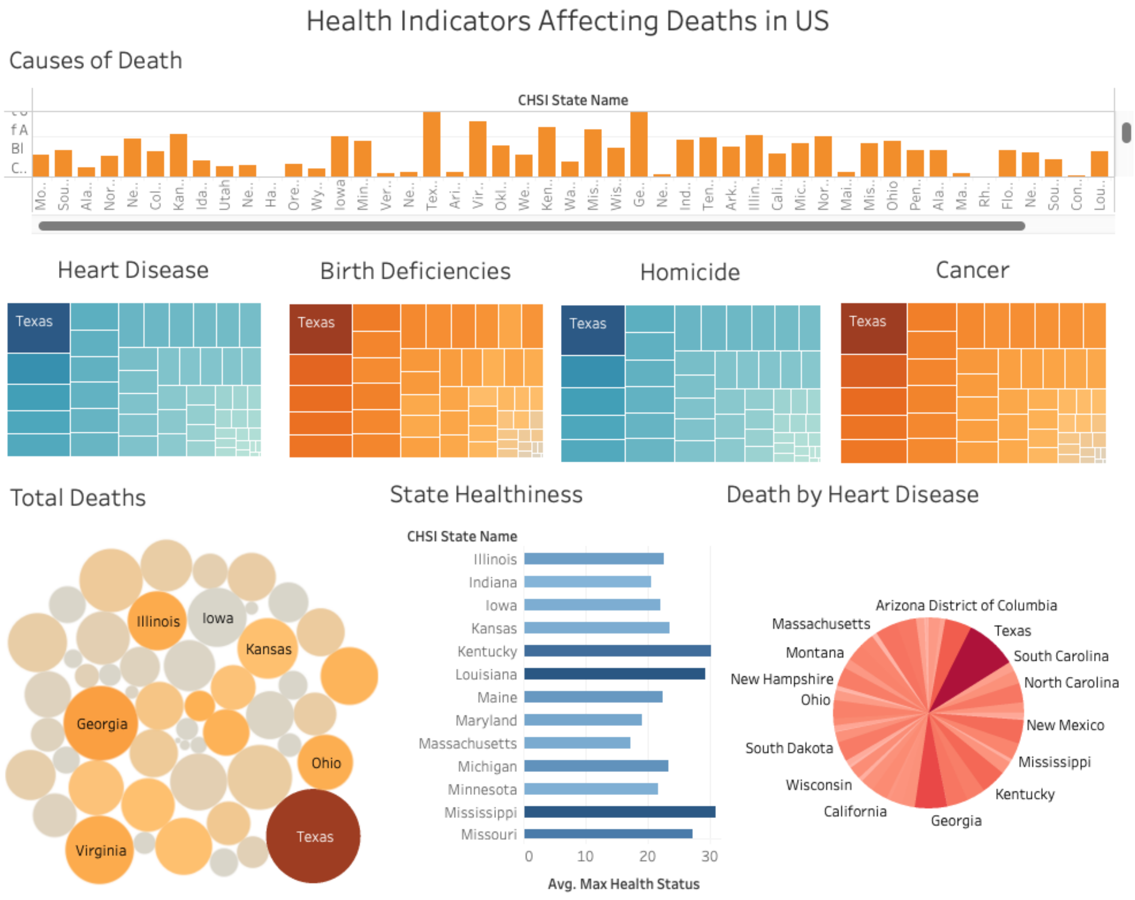
Average of Max Health Status for each CHSI State Name. Color shows average of Min Health Status.

State Healthiness

Most Healthy – Mississippi
Most Unhealthy – Texas



Total Deaths by Cancer in different States
Highest in Texas
Lowest in District of Columbia



Overall Interactive Dashboard