**File Name:** DCI Components v7.csv

**Tableau File Names:** DCIgeneral.twb, DCIhispanics.twb, DCItopcities.twb, and DCIMap.twb

**Number of Observations:** 26,125

**Number of Attributes:** 49

**Demographics (%’s of White, Hispanic, Black, Asian, and Mixed or Other Races):**

1. Open “DCI Components v5.csv” in Excel. [**Note:** “DCI Components v\*.csv” can be any file with \* being a 5 or earlier]
2. Insert 4 new columns right after “Decile\_Label”: “White”, “Hispanic”, “Black”, and “Asian”.
3. Insert a new column between “Zip\_Code” and “City”.
4. Copy the header from cell A1 for cell B1.
5. Right-click Column B and choose “Format Cells…” and “Text” before clicking on “OK.”
6. In cell B2, use the formula “=RIGHT(A2,LEN(A2)-1)”
7. Copy and paste the formula for cells B3 to B26126.
8. Insert a column between “Decile\_Label” and “White”.
9. Copy Column B, right click on new column, and choose the clipboard 123 in “Paste Options.”
10. Delete Column B.
11. Insert a new column between “Zip\_Code” and “City”.
12. Copy the header from cell A1 for cell B1.
13. Right-click Column B and choose “Format Cells…” and “Text” before clicking on “OK.”
14. In cell B2, use the formula “=CONCATENATE(A2,",")”
15. Copy and paste the formula for cells B3 to B26126.
16. Copy Column B, right click on the newly made column as described in Steps 8-9, and choose the clipboard 123 in “Paste Options.”
17. Delete Column B.
18. Copy and paste 125 entries from the newly made column as described in Step 16 onto the blank box between “Zip Code:” and “Basic Search” in the website link: [https://zipwho.com/?zip=&city=&filters=WhitePercent-0.0-100.0\_BlackPercent-0.0-100.0\_AsianPercent-0.0-100.0\_HispanicEthnicityPercent-0.0-100.0&state=AZ&mode=demo#](https://zipwho.com/?zip=&city=&filters=WhitePercent-0.0-100.0_BlackPercent-0.0-100.0_AsianPercent-0.0-100.0_HispanicEthnicityPercent-0.0-100.0&state=AZ&mode=demo)
19. When the search result appear on the right, open the Notepad and copy and paste the results from the website to Notepad. [**Note:** The first time you do it, include the headers. Then, don’t do it again. It should look like this:

Location White (%) ▾ Black (%) Asian (%) Hispanic Ethnicity (%)

1. 01225

Cheshire, MA 99.3 0.0 0.0 0.6

2. 01452

Hubbardston, MA 99.0 0.0 0.0 0.9

Then the subsequent entries after the 1st round (that is, the next block of 125 entries) will look like this:

7. 01011

Chester, MA 98.2 0.2 0.5 0.5

8. 01096

Williamsburg, MA 98.1 0.3 0.0 0.7

9. 01081

Wales, MA 98.1 0.3 0.0 0.4

10. 01068

Oakham, MA 98.0 0.1 0.1 0.8

11. 01036

Hampden, MA 97.9 0.2 0.2 0.4

]

1. Keep copying and pasting 125 entries from Excel to the website link, clicking on the “Basic Search” box, and copying and pasting the search results from the website link to Notepad until you have completed the process for all zip codes.
2. Save the Notepad file as “ZipCode.txt”.
3. Open Excel.
4. Go to “Data,” then “Get External Data,” and “From Text.”
5. Click on “ZipCode.txt,”
6. Check off “My Data has headers” and “Next.”
7. “Finish.”
8. Insert a column between Columns A and B.
9. Notice that rows 2, 4, 6, etc., contain zip codes in Column B.
10. Copy and paste from cell C2 to the very last cell of Column C onto cell B3.
11. Relabel headers “White (%) ?”, “Black (%)”, “Asian (%)”, and “Hispanic Ethnicity (%)” as “White”, “Black”, “Asian”, and “Hispanic,” respectively.
12. In Column G, label cell G1 as “Logic.”
13. Alternatively, enter “0” for even rows and “1” for odd rows beginning in cell G2 until the end of the data.
14. Click on cell G1 and then “Home,” “Sort & Filter,” and “Filter.”
15. In cell G1, click on the down button and unclick “0” and “Blanks” which will leave only rows with “1.”
16. Copy and paste the result into a new Excel File and save it as “ZipCode.csv”.
17. Delete column A.
18. Copy and paste the header from Column A for Column B.
19. In cell B2, create a formula “=CONCATENATE("Z00",A2)” and then copy and paste it for cells B3 to B???. [**Note:** It depends on how many 3-number zip codes there are to determine B???]
20. In cell B???, create a formula “=CONCATENATE("Z0",A???)” and then copy and paste it for cells B??? to B???. [**Note:** It depends on how many 4-number zip codes there are to determine B???]
21. In cell B???, create a formula “=CONCATENATE("Z",A???)” and then copy and paste it for cells B??? to B???. [**Note:** It depends on how many 5-number zip codes there are to determine B???]
22. Copy Column B, right click on Cell A1, and choose the clipboard 123 in “Paste Options.”
23. Delete Column B.
24. In R, the following code will merge “ZipCode.csv” and **“**DCI Components v\*.csv” by right inner join on “Zip\_Code” and where \* is the version number of the latter file being 5 or earlier. The final result should be “DCI Components v6.csv” with 26,125 zip codes which will be output in the folder:

merge <- read.csv(file = "DCI Components v5.csv", header = TRUE)

merge4 <- read.csv(file = "ZipCode.csv", header = TRUE)

merge5 <- merge(x = merge, y = merge4, by = "Zip\_Code", all.x = TRUE)

write.table(x = merge5, file = "DCI Components v6.csv", sep = ",", na = "", quote = FALSE, row.names = FALSE)

1. When you open "DCI Components v6.csv" in Excel, you will notice that there are missing data for demographics for the remainder of zip codes. To remedy the problem is to go to <https://statisticalatlas.com/zip/68752/Race-and-Ethnicity> where 68752 is any zip code and type in the numbers per demographic.
2. When you have completed all of the work for 26,125 zip codes, create a new column “Mixed\_Other” and the formula for each cell is 100 – White – Hispanic – Black – Asian. If the result is negative, set it as 0.
3. Create five columns in Excel called “White\_Class”, “Hispanic\_Class”, “Black\_Class”, “Asian\_Class”, and “Mixed\_Other\_Class”, and the formula for each column is “=IF(P2<25,"0-25%",IF(P2<50,"25-50%",IF(P2<75,"50-75%","75-100%")))”
4. Make sure to delete the column that has zip codes with a comma.

**Regions and SubRegions Identification by State:** <https://en.wikipedia.org/wiki/List_of_regions_of_the_United_States>

**Northeast:**

**New England:** CT, ME, MA, NH, RI, VT

**Mid-Atlantic:** NJ, NY, PA

**Midwest:**

**East North Central:** IL, IN, MI, OH, WI

**West North Central:** IA, KS, MN, MO, NE, ND, SD

**South:**

**South Atlantic:** DE, FL, GA, MD, NC, SC, VA, DC, WV

**East South Central:** AL, KY, MS, TN

**West South Central:** AR, LA, OK, TX

**West:**

**Mountain:** AZ, CO, ID, MT, NV, NM, UT, WY

**Pacific:** AK, CA, HI, OR, WA

1. The above is the reference of how to label each zip code based on what state it is located in.
2. In R, the following code will create 2 columns for Region and SubRegion to correspond with the zip codes:

merge1 <- read.csv(file = "DCI Components v6.csv", header = TRUE)

merge1$Region[merge1$State == "CT"] <- "Northeast"

merge1$Region[merge1$State == "ME"] <- "Northeast"

merge1$Region[merge1$State == "MA"] <- "Northeast"

merge1$Region[merge1$State == "NH"] <- "Northeast"

merge1$Region[merge1$State == "RI"] <- "Northeast"

merge1$Region[merge1$State == "VT"] <- "Northeast"

merge1$Region[merge1$State == "NJ"] <- "Northeast"

merge1$Region[merge1$State == "NY"] <- "Northeast"

merge1$Region[merge1$State == "PA"] <- "Northeast"

merge1$Region[merge1$State == "IL"] <- "Midwest"

merge1$Region[merge1$State == "IN"] <- "Midwest"

merge1$Region[merge1$State == "MI"] <- "Midwest"

merge1$Region[merge1$State == "OH"] <- "Midwest"

merge1$Region[merge1$State == "WI"] <- "Midwest"

merge1$Region[merge1$State == "IA"] <- "Midwest"

merge1$Region[merge1$State == "KS"] <- "Midwest"

merge1$Region[merge1$State == "MN"] <- "Midwest"

merge1$Region[merge1$State == "MO"] <- "Midwest"

merge1$Region[merge1$State == "NE"] <- "Midwest"

merge1$Region[merge1$State == "ND"] <- "Midwest"

merge1$Region[merge1$State == "SD"] <- "Midwest"

merge1$Region[merge1$State == "DE"] <- "South"

merge1$Region[merge1$State == "FL"] <- "South"

merge1$Region[merge1$State == "GA"] <- "South"

merge1$Region[merge1$State == "MD"] <- "South"

merge1$Region[merge1$State == "NC"] <- "South"

merge1$Region[merge1$State == "SC"] <- "South"

merge1$Region[merge1$State == "VA"] <- "South"

merge1$Region[merge1$State == "DC"] <- "South"

merge1$Region[merge1$State == "WV"] <- "South"

merge1$Region[merge1$State == "AL"] <- "South"

merge1$Region[merge1$State == "KY"] <- "South"

merge1$Region[merge1$State == "MS"] <- "South"

merge1$Region[merge1$State == "TN"] <- "South"

merge1$Region[merge1$State == "AR"] <- "South"

merge1$Region[merge1$State == "LA"] <- "South"

merge1$Region[merge1$State == "OK"] <- "South"

merge1$Region[merge1$State == "TX"] <- "South"

merge1$Region[merge1$State == "AZ"] <- "West"

merge1$Region[merge1$State == "CO"] <- "West"

merge1$Region[merge1$State == "ID"] <- "West"

merge1$Region[merge1$State == "MT"] <- "West"

merge1$Region[merge1$State == "NV"] <- "West"

merge1$Region[merge1$State == "NM"] <- "West"

merge1$Region[merge1$State == "UT"] <- "West"

merge1$Region[merge1$State == "WY"] <- "West"

merge1$Region[merge1$State == "AK"] <- "West"

merge1$Region[merge1$State == "CA"] <- "West"

merge1$Region[merge1$State == "HI"] <- "West"

merge1$Region[merge1$State == "OR"] <- "West"

merge1$Region[merge1$State == "WA"] <- "West"

merge1$Region[merge1$State == "CT"] <- "New England"

merge1$SubRegion[merge1$State == "ME"] <- "New England"

merge1$SubRegion[merge1$State == "MA"] <- "New England"

merge1$SubRegion[merge1$State == "NH"] <- "New England"

merge1$SubRegion[merge1$State == "RI"] <- "New England"

merge1$SubRegion[merge1$State == "VT"] <- "New England"

merge1$SubRegion[merge1$State == "NJ"] <- "Mid-Atlantic"

merge1$SubRegion[merge1$State == "NY"] <- "Mid-Atlantic"

merge1$SubRegion[merge1$State == "PA"] <- "Mid-Atlantic"

merge1$SubRegion[merge1$State == "IL"] <- "East North Central"

merge1$SubRegion[merge1$State == "IN"] <- "East North Central"

merge1$SubRegion[merge1$State == "MI"] <- "East North Central"

merge1$SubRegion[merge1$State == "OH"] <- "East North Central"

merge1$SubRegion[merge1$State == "WI"] <- "East North Central"

merge1$SubRegion[merge1$State == "IA"] <- "West North Central"

merge1$SubRegion[merge1$State == "KS"] <- "West North Central"

merge1$SubRegion[merge1$State == "MN"] <- "West North Central"

merge1$SubRegion[merge1$State == "MO"] <- "West North Central"

merge1$SubRegion[merge1$State == "NE"] <- "West North Central"

merge1$SubRegion[merge1$State == "ND"] <- "West North Central"

merge1$SubRegion[merge1$State == "SD"] <- "West North Central"

merge1$SubRegion[merge1$State == "DE"] <- "South Atlantic"

merge1$SubRegion[merge1$State == "FL"] <- "South Atlantic"

merge1$SubRegion[merge1$State == "GA"] <- "South Atlantic"

merge1$SubRegion[merge1$State == "MD"] <- "South Atlantic"

merge1$SubRegion[merge1$State == "NC"] <- "South Atlantic"

merge1$SubRegion[merge1$State == "SC"] <- "South Atlantic"

merge1$SubRegion[merge1$State == "VA"] <- "South Atlantic"

merge1$SubRegion[merge1$State == "DC"] <- "South Atlantic"

merge1$SubRegion[merge1$State == "WV"] <- "South Atlantic"

merge1$SubRegion[merge1$State == "AL"] <- "East South Central"

merge1$SubRegion[merge1$State == "KY"] <- "East South Central"

merge1$SubRegion[merge1$State == "MS"] <- "East South Central"

merge1$SubRegion[merge1$State == "TN"] <- "East South Central"

merge1$SubRegion[merge1$State == "AR"] <- "West South Central"

merge1$SubRegion[merge1$State == "LA"] <- "West South Central"

merge1$SubRegion[merge1$State == "OK"] <- "West South Central"

merge1$SubRegion[merge1$State == "TX"] <- "West South Central"

merge1$SubRegion[merge1$State == "AZ"] <- "Mountain"

merge1$SubRegion[merge1$State == "CO"] <- "Mountain"

merge1$SubRegion[merge1$State == "ID"] <- "Mountain"

merge1$SubRegion[merge1$State == "MT"] <- "Mountain"

merge1$SubRegion[merge1$State == "NV"] <- "Mountain"

merge1$SubRegion[merge1$State == "NM"] <- "Mountain"

merge1$SubRegion[merge1$State == "UT"] <- "Mountain"

merge1$SubRegion[merge1$State == "WY"] <- "Mountain"

merge1$SubRegion[merge1$State == "AK"] <- "Pacific"

merge1$SubRegion[merge1$State == "CA"] <- "Pacific"

merge1$SubRegion[merge1$State == "HI"] <- "Pacific"

merge1$SubRegion[merge1$State == "OR"] <- "Pacific"

merge1$SubRegion[merge1$State == "WA"] <- "Pacific"

merge1 <- merge1[ , c(1:3, 26:27, 4:25)]

write.table(x = merge1, file = " DCI Components v7.csv", sep = ",", na = "", quote = FALSE, row.names = FALSE)

**Area Square Miles for Land or Water, Longitude, and Latitude:**

1. Go to <https://www.census.gov/geo/maps-data/data/gazetteer2017.html>.
2. Scroll down until you see “ZIP Code Tabulation Areas.”
3. Click on “Download the ZIP Code Tabulation Areas Gazetteer File (976KB).”
4. Save the file, 2017\_Gaz\_zcta\_national.txt, in a folder on Desktop.

|  |  |  |
| --- | --- | --- |
| **Column** | **Label** | **Description** |
| Column 1 | GEOID | Five digit ZIP Code Tabulation Area Census Code |
| Column 2 | ALAND | Land Area (square meters) - Created for statistical purposes only |
| Column 3 | AWATER | Water Area (square meters) - Created for statistical purposes only |
| Column 4 | ALAND\_SQMI | Land Area (square miles) - Created for statistical purposes only |
| Column 5 | AWATER\_SQMI | Water Area (square miles) - Created for statistical purposes only |
| Column 6 | INTPTLAT | Latitude (decimal degrees) First character is blank or "-" denoting North or South latitude respectively |
| Column 7 | INTPTLONG | Longitude (decimal degrees) First character is blank or "-" denoting East or West longitude respectively |

1. Open Excel.
2. Go to “Data,” then “Get External Data,” and “From Text.”
3. Click on “2017\_Gaz\_zcta\_national.txt,”
4. Check off “My Data has headers” and “Next.”
5. “Finish”
6. Rename “GEOID” as “Zip\_Code”, “INTPLAT” as “LATITUDE”, and “INTPTLONG” as “LONGITUDE”.
7. Delete the following columns: “ALAND” and “AWATER”.
8. Insert a column between “Zip\_Code” and “ALAND\_SQMI.”
9. Copy and paste the header from Column A for Column B.
10. In cell B2, create a formula “=CONCATENATE("Z00",A2)” and then copy and paste it for cells B3 to B143.
11. In cell B144, create a formula “=CONCATENATE("Z0",A144)” and then copy and paste it for cells B145 to B2570.
12. In cell B2571, create a formula “=CONCATENATE("Z",A2571)” and then copy and paste it for cells B2572 to B33145.
13. Copy Column B, right click on Cell A1, and choose the clipboard 123 in “Paste Options.”
14. Delete Column B.
15. Save the file as “gazetteer.csv”.
16. In R, the following code will merge “gazetteer.csv” and an older version of **“**DCI Components v7.csv” by left inner join on “Zip\_Code”. The final result should be “DCI Components v7.csv” with 26,125 zip codes which will be output in the folder:

merge2 <- read.csv(file = "newdata1.csv", header = TRUE)

merge3 <- merge(x = merge1, y = merge2, by = "Zip\_Code", all.x = TRUE)

write.table(x = merge3, file = " DCI Components v7.csv", sep = ",", na = "", quote = FALSE, row.names = FALSE)

**Criteria for classifying population density by land area in square miles**

=IF(AJ2<25,"0-25",IF(AJ2<50,"25-50",IF(AJ2<75,"50-75",IF(AJ2<100,"75-100",IF(AJ2<250,"100-250",IF(AJ2<500,"250-500",IF(AJ2<750,"500-750",IF(AJ2<1000,"750-1000",IF(AJ2<2500,"1000-2500",IF(AJ2<5000,"2500-5000",IF(AJ2<7500,"5000-7500",IF(AJ2<10000,"7500-10000","10000+"))))))))))))

**Criteria for classifying percentage of No Diploma**

=IF(G2<5,"0-5%",IF(G2<10,"5-10%",IF(G2<15,"10-15%",IF(G2<20,"15-20%",IF(G2<25,"20-25%",IF(G2<30,"25-30%","30+%"))))))

**Criteria for classifying percentage of Vacancy**

=IF(J2<5,"0-5%",IF(J2<10,"5-10%",IF(J2<15,"10-15%",IF(J2<20,"15-20%","20+%"))))

**Criteria for classifying percentage of No Job**

=IF(M2<10,"0-10%",IF(M2<20,"10-20%",IF(M2<30,"20-30%",IF(M2<40,"30-40%",IF(M2<50,"40-50%","50+%")))))

**Criteria for classifying percentage of Poverty**

=IF(P2<5,"0-5%",IF(P2<10,"5-10%",IF(P2<15,"10-15%",IF(P2<20,"15-20%",IF(P2<25,"20-25%",IF(P2<30,"25-30%","30+%"))))))

**Criteria for classifying percentage of Med Inc**

=IF(S2<30,"0-30%",IF(S2<60,"30-60%",IF(S2<90,"60-90%",IF(S2<120,"90-120%",IF(S2<150,"120-150%","150+%")))))

**Criteria for classifying percentage of Job Chg**

=IF(V2<-30,"-30-%",IF(V2<-20,"-30--20%",IF(V2<-10,"-20--10%",IF(V2<0,"-10-0%",IF(V2<10,"0-10%",IF(V2<20,"10-20%",IF(V2<30,"20-30%","30+%")))))))

**Criteria for classifying percentage of Chg Biz**

=IF(V2<-30,"-30-%",IF(V2<-20,"-30--20%",IF(V2<-10,"-20--10%",IF(V2<0,"-10-0%",IF(V2<10,"0-10%",IF(V2<20,"10-20%",IF(V2<30,"20-30%","30+%")))))))

**Aggregated Results:** The purpose of the following code in R is to combine all zip codes per City per State by getting the average of selected variables:

dataset <- read.csv(file = "DCI Components v7.csv", header = TRUE)

df2 <- count(dataset, City, State)

library(dplyr)

df1 <- group\_by(dataset, City, State) %>%

summarize(Sum\_Population = sum(Population),

Mean\_No\_Diploma = mean(No\_Diploma),

Mean\_Vacancy = mean(Vacancy),

Mean\_No\_Job = mean(No\_Job),

Mean\_Poverty = mean(Poverty),

Mean\_Med\_Inc = mean(Med\_Inc),

Mean\_Chg\_Job = mean(Chg\_Job),

Mean\_Chg\_Biz = mean(Chg\_Biz),

Mean\_DCI\_Percentile = mean(DCI\_Percentile),

Mean\_White = mean(White),

Mean\_Hispanic = mean(Hispanic),

Mean\_Black = mean(Black),

Mean\_Asian = mean(Asian),

Mean\_Mixed\_Other = mean(Mixed\_Other),

Sum\_Area\_Land = sum(Area\_Land),

Sum\_Area\_Water = sum(Area\_Water)

)

merge14 <- merge(x = df2, y = df1, by = c("City", "State"), all = TRUE)

names(merge14)[names(merge14) == "n"] <- 'No\_of\_Zip\_Codes'

write.table(x = merge14, file = "DCI Components Aggregated v1.csv", sep = ",", na = "", quote = FALSE, row.names = FALSE)

**Tableau:**

**All Zip Codes:** The Tableau file for graphs is DCIgeneral.twb and is based on DCI Components v7.csv.

**Top 27 Cities:** Top 27 Cities.csv is based on cities with at least 30 zip codes which came from DCI Components Aggregated v1.csv. The Tableau file for graphs is DCI27cities.twb.

**Similar Cities to San Antonio:** Hispanic Cities.csv is based on cities that have between 35 and 65% Hispanics, inclusively, and a population of at least 250,000 which came from DCI Components Aggregated v1.csv. The Tableau file for graphs is DCIhispanics.twb.

**Maps of San Antonio and Nation:** The Tableau file for graphs is DCIMap.twb and is based on DCI Components v7.csv.