Spatial Analysis of Charlotte Transit Frequency

Julia Bright

2025-04-10

Code can be found at: https://github.com/JBrighttt/plan372_hmks/tree/main/Final%20Project

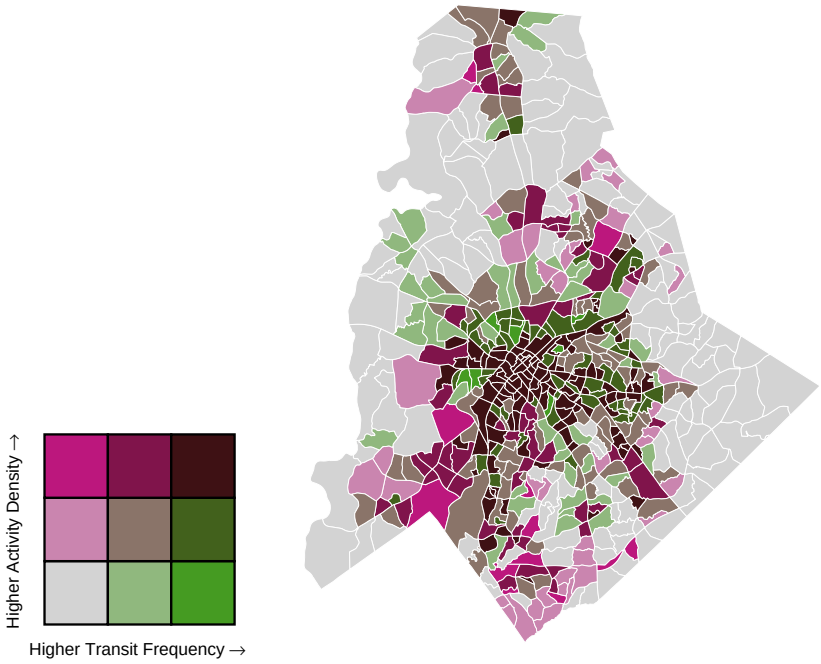
```
#read in the data
median_income <-
↪  st_read("/Users/julia/Library/CloudStorage/OneDrive-UniversityofNorthCarolinaatChapelHill/Documents/
↪  Project/median
↪  income/SimplyAnalytics_Shapefiles_d1a362ad9f6aa354772ce38e8f811dd5e0f3080d168c7c1ec9cff332fac4bcc5.s
#name variables
median_income <- median_income %>%
  rename(FIPS = spatial_id,
         income = VALUE0) %>%
  select(FIPS, income, #Average household yearly income by census block group - ACS
   ↪  Five-Year Estimates 2022
         geometry)

#read in data
transport_costs <-
↪  st_read("/Users/julia/Library/CloudStorage/OneDrive-UniversityofNorthCarolinaatChapelHill/Documents/
↪  Project/transportation
↪  costs/SimplyAnalytics_Shapefiles_212f6759e776edf4c0b452bc173d29e4410b09e0b0882cd3dfe002ac631f23b6.sl
#name variables
transport_costs <- transport_costs %>%
  rename(FIPS = spatial_id,
         transport_costs = VALUE0) %>%
  select(FIPS, transport_costs, #Household yearly average spending on transportation by
   ↪  census block group - Consumer Expenditure Estimates, using CE PUMD and ACS
   ↪  Five-Year Estimates 2022
         geometry)

SLD <-
↪  st_read("/Users/julia/Library/CloudStorage/OneDrive-UniversityofNorthCarolinaatChapelHill/Documents/
↪  Project/SmartLocationDatabaseV3/SmartLocationDatabase.gdb")

#filter to mecklenburg county
SLD <- SLD %>%
  filter(STATEFP == 37) %>%
  filter(COUNTYFP == 119)

#clean dataframe to specific variables
```

```
SLD <- SLD %>%
  select(GEOID20, CBSA_Name,
         R_LowWageWk, # Count of workers earning $1250/month or less (home location) -
↪  2017 Census LEHD RAC
         R_PCTLOWWAGE, # Percent of low wage workers in a CBG (home location) - 2017
↪  Census LEHD RAC
         E_LowWageWk, # Number of workers earning $1250/month or less (work location) -
↪  2017 Census LEHD WAC
         E_PctLowWage, # Percent LowWageWk of total number of workers in a CBG (work
↪  location) - 2017 Census LEHD WAC
         D1C, # Gross employment density (jobs/acre) on unprotected land - ACS Five-Year
↪  Estimates (2014-2018), refined with other SLD variables
         D1D, # Gross activity density (employment + HUs) on unprotected land - ACS
↪  Five-Year Estimates (2014-2018), refined with other SLD variables
         D4D # Aggregate frequency of transit service [D4c] per square mile -  Derived
↪  from other SLD variables
         )

#naming variables
SLD <- SLD %>%
  rename(FIPS = GEOID20,
         geometry=Shape,
         CountLowWage_home=R_LowWageWk,
         PctLowWage_home=R_PCTLOWWAGE,
         CountLowWage_work=E_LowWageWk,
         PctLowWage_work=E_PctLowWage,
         JobsPerAcre=D1C,
         GrossActivityDensity=D1D,
         TransitFrequency=D4D)
```

```
# reproject SLD to match clt transit costs
SLD <- SLD %>%
  st_transform(4326)

#make geometries valid
median_income <- st_make_valid(median_income)
transport_costs <- st_make_valid(transport_costs)
SLD <- st_make_valid(SLD)

#join the datasets
clt_costs_income <- left_join(transport_costs, median_income %>%
                    as.data.frame() %>%
                    select(-geometry))
#final dataset
data <- left_join(SLD, clt_costs_income %>%
                    as.data.frame() %>%
                    select(-geometry))

#recode the variable that means no transit service in the CBG to 0
data <- data %>%
  mutate(TransitFrequency = case_when(TransitFrequency==-99999 ~ 0,
                                      TRUE ~ TransitFrequency))
```

Transit Frequency and Gross Activity Density

```
# credit to Chris Prener for biscale ggplot tutorials
↪  https://chris-prener.github.io/biscale/articles/biscale.html#bivariate-mapping-with-biscale
↪

#create bivariate map of transit freq and activity density

#step 1 create class for the variables
data1 <- bi_class(data, x=TransitFrequency, y=GrossActivityDensity, style = "quantile",
↪  dim=3)

#make the map
map1 <- ggplot() +
  geom_sf(data = data1, mapping = aes(fill = bi_class), color = "white", size = 0.1,
  ↪  show.legend = FALSE) +
  bi_scale_fill(pal = "PinkGrn", dim = 3) +
  theme_void()

#make the legend
legend1 <- bi_legend(pal = "PinkGrn",
                     pad_width = 0.5,
                     pad_color = "black",
                     dim = 3,
                     xlab = "Higher Transit Frequency",
                     ylab = "Higher Activity Density",
                     size = 8)

#add map and legend together
finalmap1 <- ggdraw() +
  draw_plot(map1, 0, 0, 1, 1) +
  draw_plot(legend1, -0.05, 0, .4, .4)
finalmap1
```



Area of Northlake Mall with medium activity density, mostly due to jobs, but low transit frequency. Likewise with Concord Mills in the Northeast. Southeast area near Matthews also high activity density and medium transit frequency. Southern Charlotte the same, but this area is less commercial and more high-income residential.

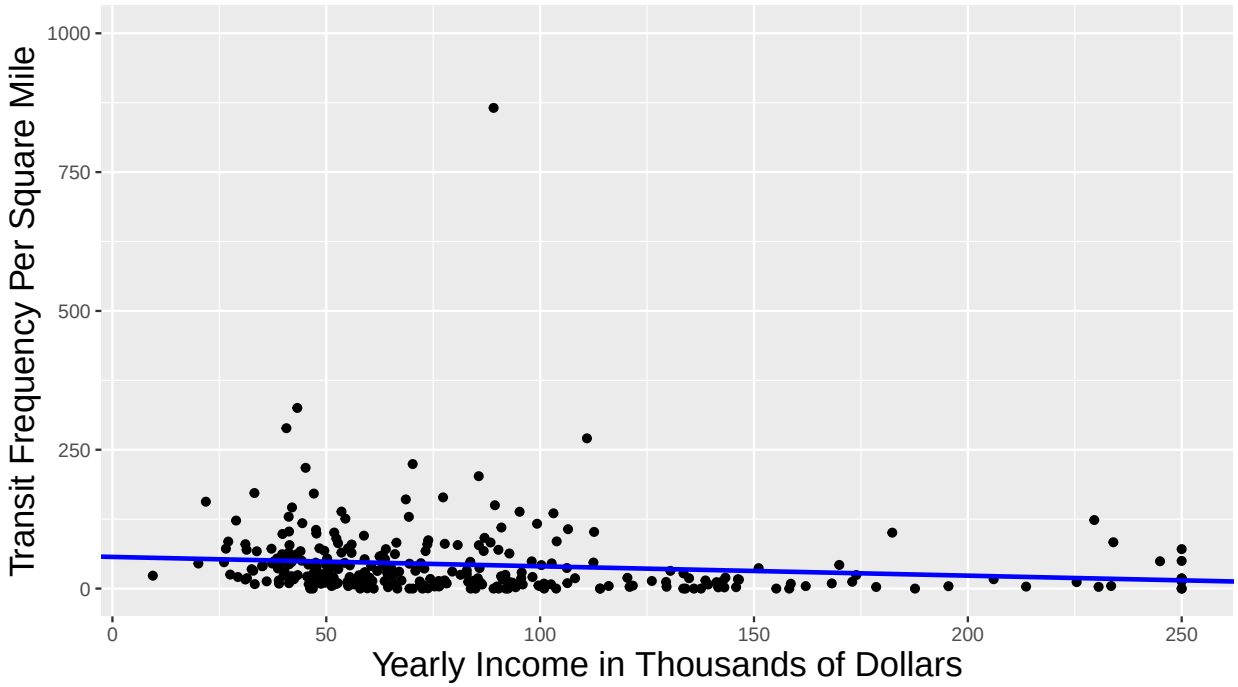
```
#regression to analyze relationships between income and transit service

#create segmented income variable to better analyze the relationship
data$income_thousands <- data$income / 1000

fit1 <- lm(TransitFrequency ~ income_thousands, data=data)
summary(fit1)
```

```
##
## Call:
## lm(formula = TransitFrequency ~ income_thousands, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -49.20 -32.85 -17.81   8.83 823.71
##
## Coefficients:
##                  Estimate Std. Error t value          Pr(>|t|)
## (Intercept)      56.99550    7.32609   7.780 0.000000000000116 ***
## income_thousands -0.16863    0.07478  -2.255            0.0248 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 67.29 on 302 degrees of freedom
```

```
##   (251 observations deleted due to missingness)
## Multiple R-squared:  0.01656,    Adjusted R-squared:  0.0133
## F-statistic: 5.085 on 1 and 302 DF,  p-value: 0.02484
```

```r
summary(data$transport_costs)
summary(data$income_thousands) #THE INCOME IS CAPPED AT 250K THIS EXPLAINS WHY THE
↪  REGRESSION DOESN'T SHOW THE RELATIONSHIP I WOULD EXPECT
length(data$income_thousands == 250.001) #There are 555 entries that show 250k income but
↪  highly likely that many of them are substantially higher
```

```r
#plot the data and trend line

ggplot(data = data, aes(x=income_thousands, y=TransitFrequency)) +
  geom_point() +
  geom_abline(intercept = 56.99550, slope = -0.16863, size=1, color="blue") +
  coord_cartesian(ylim = c(0,1000)) +
  labs(x="Yearly Income in Thousands of Dollars", y="Transit Frequency Per Square Mile")
  ↪  +
  theme(axis.title = element_text(size=16))
```




```r
#regression to analyze relationship between % income spent on transportation and transit
↪  service

#create percentage income spent on transport variable
data$PctTransportCosts <- (data$transport_costs / data$income) * 100

#remove outlier which seems to be faulty data
data$PctTransportCosts[data$PctTransportCosts > 100] = NA

fit2 <- lm(TransitFrequency ~ PctTransportCosts, data=data)
```

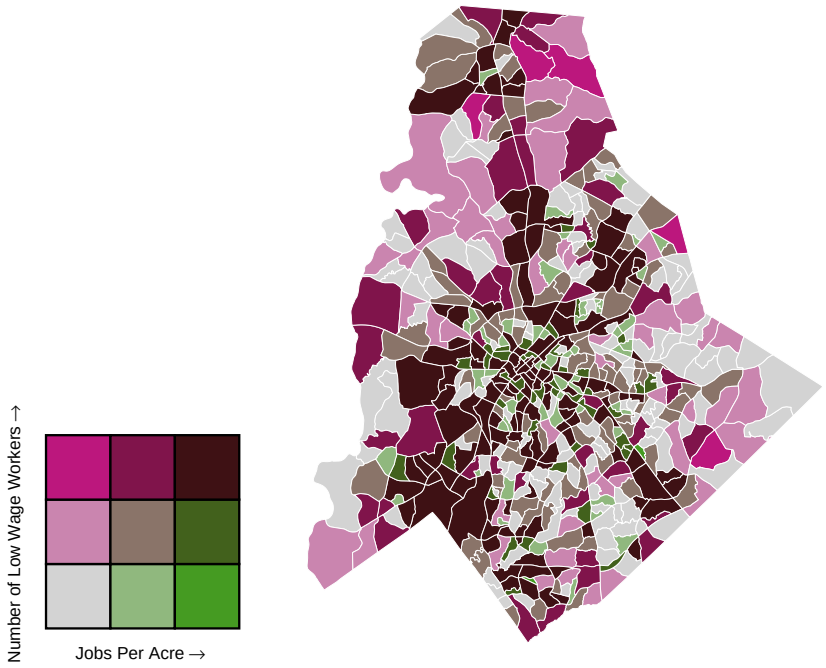
```r
summary(fit2)
```

```
##
## Call:
## lm(formula = TransitFrequency ~ PctTransportCosts, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -52.18 -31.78 -19.96  10.52 829.71
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        21.9796    11.2612   1.952   0.0519 .
## PctTransportCosts   0.9838     0.4944   1.990   0.0475 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 67.51 on 301 degrees of freedom
##   (252 observations deleted due to missingness)
## Multiple R-squared:  0.01298,    Adjusted R-squared:  0.009706
## F-statistic:  3.96 on 1 and 301 DF,  p-value: 0.0475
```

```r
#plot the data and trend line
ggplot(data, aes(x=PctTransportCosts, y=TransitFrequency)) +
  geom_point() +
  geom_abline(intercept = 21.9796, slope = 0.9838, size=1, color="blue") +
  coord_cartesian(ylim = c(0,1000)) +
  labs(x="Percentage of Income Spent on Transportation", y="Transit Frequency Per Square
  ↪  Mile") +
  theme(axis.title = element_text(size=16))
```

Jobs Per Acre and Number of Low Wage Workers (Workplace)

```
#create bivariate map of jobs per acre and number of low wage workers (work location)

#create class for the variables
data2_num <- bi_class(data, x=JobsPerAcre, y=CountLowWage_work, style = "quantile",
↪  dim=3)

#make the map
map2_num <- ggplot() +
  geom_sf(data = data2_num, mapping = aes(fill = bi_class), color = "white", size = 0.1,
  ↪  show.legend = FALSE) +
  bi_scale_fill(pal = "PinkGrn", dim = 3) +
  theme_void()

#make the legend
legend2_num <- bi_legend(pal = "PinkGrn",
                    pad_width = 0.5,
                    pad_color = "black",
                    dim = 3,
                    xlab = "Jobs Per Acre",
                    ylab = "Number of Low Wage Workers",
                    size = 8)
#combine map and legend
finalmap2_num <- ggdraw() +
  draw_plot(map2_num, 0, 0, 1, 1) +
  draw_plot(legend2_num, -0.05, 0, .4, .4)
finalmap2_num
```

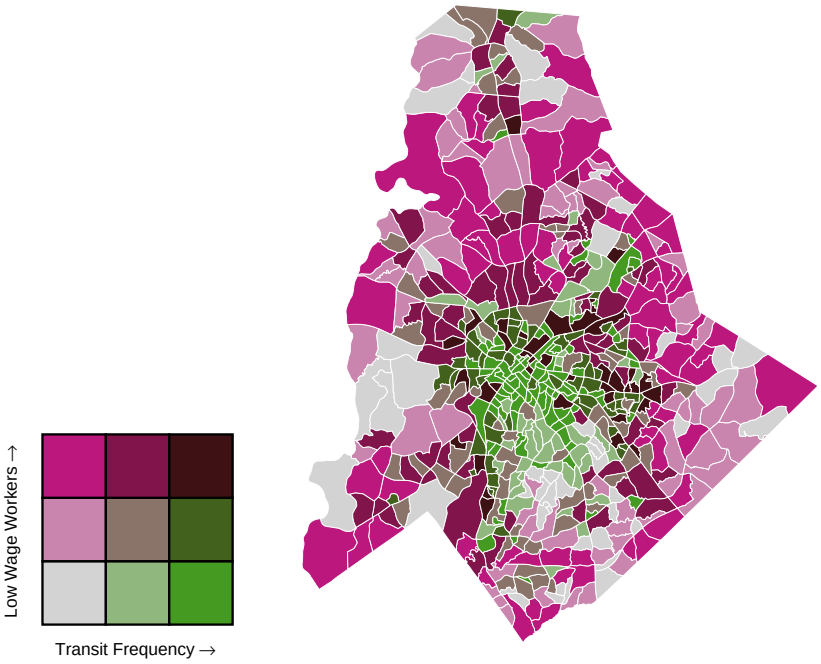


Around I-77 north of the City, large concentration of jobs and low wage jobs. In this area of several block groups is Northlake Mall to the northwest, freight and business to the southeast, and residential to the southwest. Also see pockets around N Tryon/Highway 29: Tryon Hills, Sugar Creek

Additionally, area of University City around UNC Charlotte shows large concentrations.

Transit Frequency and Number of Low Wage Workers (Home)

```
#create bivariate map of transit frequency and number of low wage workers (home location)

#create class for the variables
data3_num <- bi_class(data, x=TransitFrequency, y=CountLowWage_home, style = "quantile",
↪  dim=3)

#make the map
map3_num <- ggplot() +
  geom_sf(data = data3_num, mapping = aes(fill = bi_class), color = "white", size = 0.1,
  ↪  show.legend = FALSE) +
  bi_scale_fill(pal = "PinkGrn", dim = 3) +
  theme_void()

#make the legend
legend3_num <- bi_legend(pal = "PinkGrn",
                    pad_width = 0.5,
                    pad_color = "black",
                    dim = 3,
                    xlab = "Transit Frequency",
                    ylab = "Low Wage Workers",
                    size = 8)
#combine map and legend
finalmap3_num <- ggdraw() +
  draw_plot(map3_num, 0, 0, 1, 1) +
  draw_plot(legend3_num, -0.05, 0, .4, .4)
finalmap3_num
```



Identified an area around N Sharon Amity Rd with high number of low wage worker homes, and high of transit frequency. Also area around Highway 29/N Tryon, Neighborhoods of Hidden Valley and surrounding Shannon Park.

Around Northlake Mall once again there are high numbers of low wage workers but low transit frequency.

Areas around UNC Charlotte with many low wage workers' homes but poor transit service. Thinking up the road into Harrisburg, and southeast in surburban areas off of Highway 24/W.T. Harris Blvd.