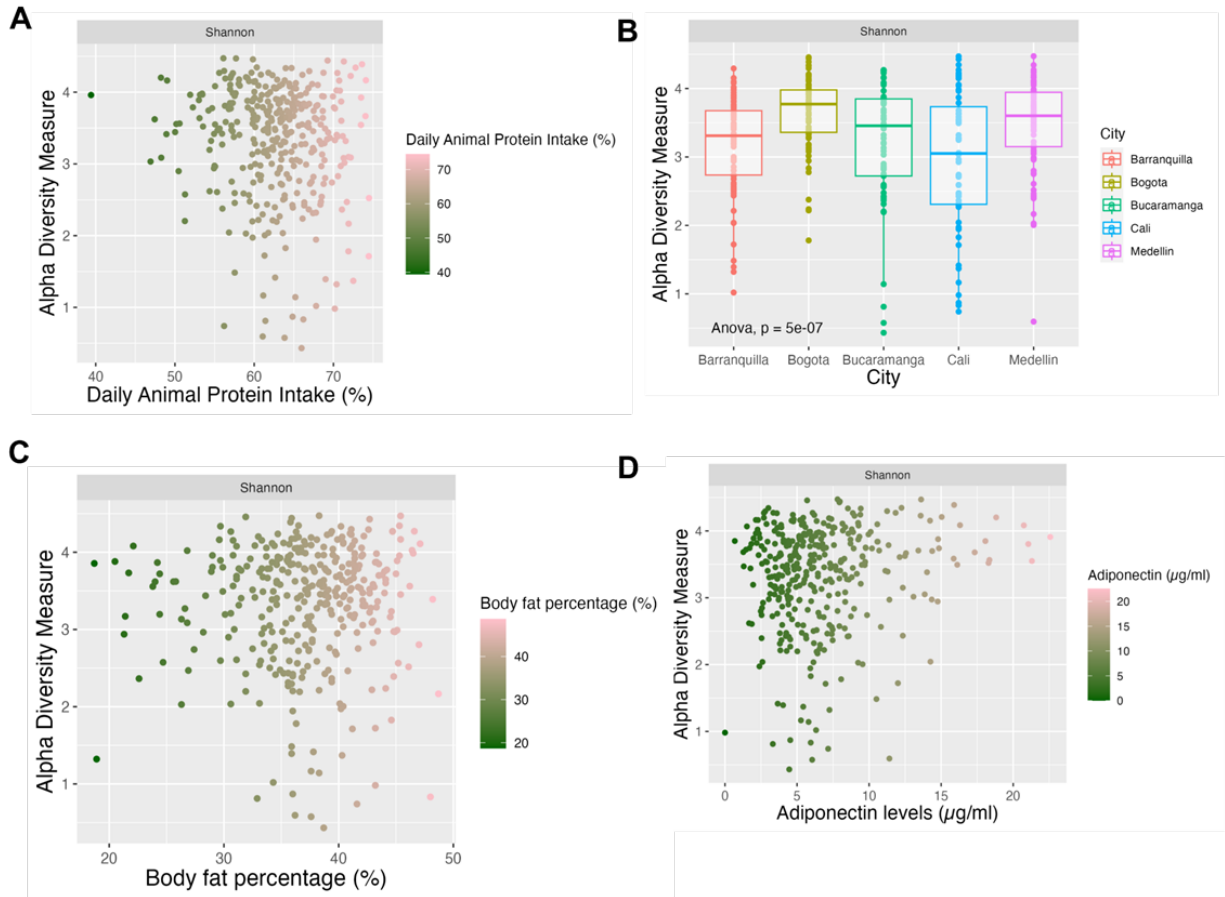
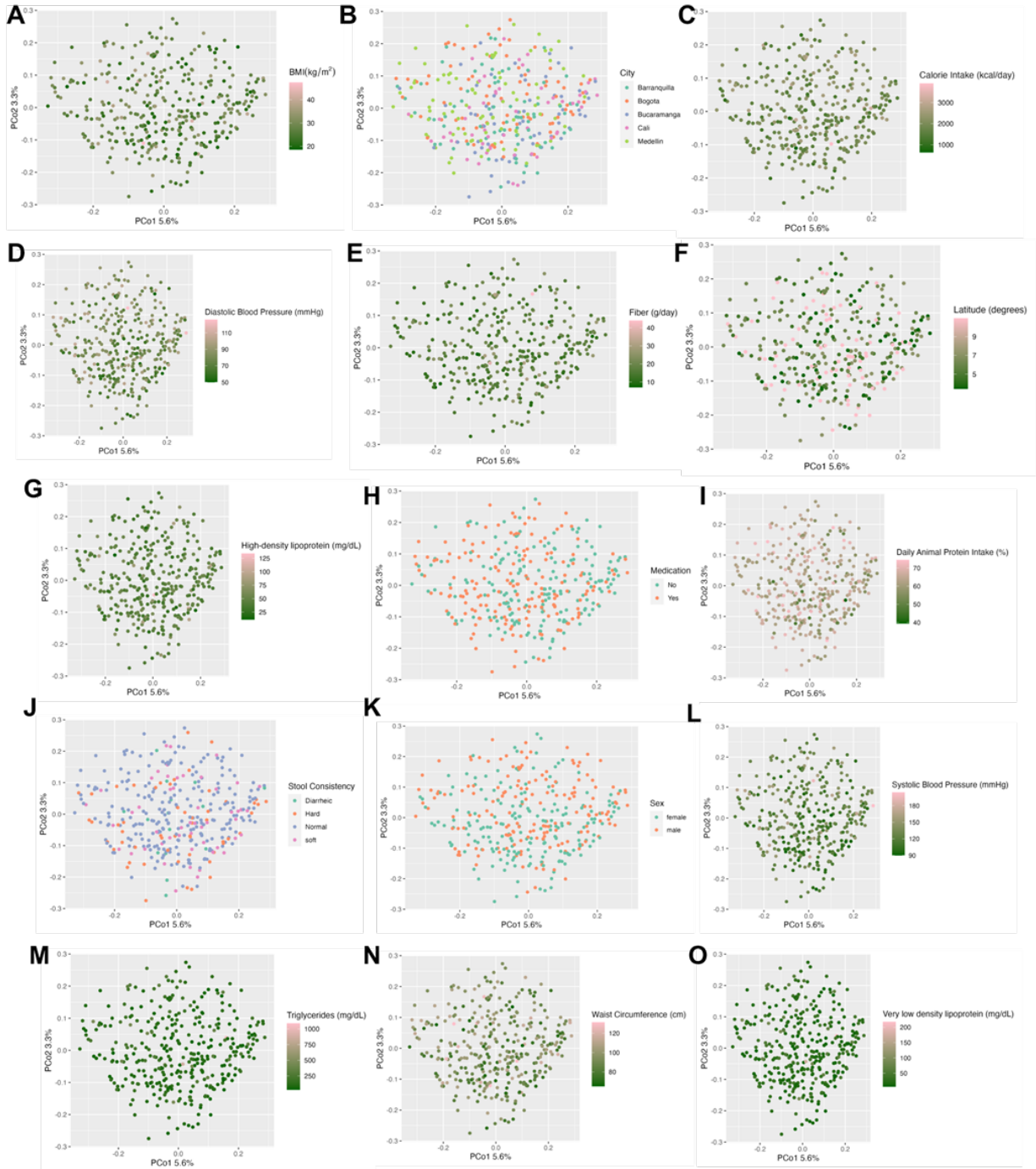


## SUPPLEMENTAL MATERIAL

R code for all project aims can be found at the following website:  
<https://github.com/sanasamadi/MICB475>.



**Figure S1. Select alpha diversity plots showed interesting results.** Alpha-diversity analysis was performed on all 40 predictors within the dataset (Shannon). Select predictors are shown above.  $n = 442$ .



**Figure S2. Select predictors were important for variation in beta-diversity across multiple aims.** Beta-diversity analysis was performed on all 40 predictors within the dataset (PERMANOVA, Unweighted Unifrac). Select predictors are shown above. n = 442.

**Table S1. Filtering out “Body Fat Percentage” from the full model caused “Insulin” and “Medication (Yes)” to be dropped in the final model.** A linear model analysis (response variable = Shannon Index) was performed on the full model of 34 predictors. Adjusted AIC (AICc) score for full model = 862.4. AICc score for the final model = -205.2. Estimates indicate the increase or decrease in Shannon diversity. For the following categorical variables: Cities are in reference to Barranquilla city, and “Sex (Male)” is in reference to “Sex (Female)”. Significant results are indicated by an asterisk (\* =  $P < 0.05$ , \*\* =  $P < 0.01$ , \*\*\* =  $P < 0.001$ , \*\*\*\* =  $P < 0.0001$ ). n = 361.

Metadata Category	Estimate	P-Value	Significance
Age (Years)	0.011	0.128	ns
City (Bogota)	0.494	0.000	****
City (Bucaramanaga)	0.029	0.843	ns
City (Cali)	-0.245	0.048	*
City (Medellin)	0.411	0.003	**
Carbohydrates (% Daily Value)	-0.075	0.017	*
Total Fat (% Daily Value)	-0.089	0.012	*

Animal Protein (% Daily Value)	-0.029	0.001	**
--------------------------------	--------	-------	----

Sex (Male)	-0.221	0.106	ns
------------	--------	-------	----

---

**Table S2. Filtering out “Insulin” from the full model caused “Body Fat Percentage” and “Medication (Yes)” to be dropped in the final model.** A linear model analysis (response variable = Shannon Index) was performed on the full model of 34 predictors. Adjusted AIC (AICc) score for full model = 860.1. AICc score for the final model = -205.2. Estimates indicate the increase or decrease in Shannon diversity. For the following categorical variables: Cities are in reference to Barranquilla city, and “Sex (Male)” is in reference to “Sex (Female)”. Significant results are indicated by an asterisk (\* =  $P < 0.05$ , \*\* =  $P < 0.01$ , \*\*\* =  $P < 0.001$ , \*\*\*\* =  $P < 0.0001$ ). n = 361.

Metadata Category	Estimate	P-Value	Significance
Age (Years)	0.011	0.128	ns
City (Bogota)	0.494	0.000	****
City (Bucaramanaga)	0.029	0.843	ns
City (Cali)	-0.245	0.048	*
City (Medellin)	0.411	0.003	**
Carbohydrates (% Daily Value)	-0.075	0.017	*
Total Fat (% Daily Value)	-0.089	0.012	*

Animal Protein (% Daily Value)	-0.029	0.001	**
Sex (Male)	-0.221	0.106	ns

---

**Table S3. Filtering out “Medication (Yes)” from the full model caused “Body Fat Percentage” and “Insulin” to be dropped in the final model.** A linear model analysis (response variable = Shannon Index) was performed on the full model of 34 predictors. Adjusted AIC (AICc) score for full model = 860.4. AICc score for the final model = -205.2. Estimates indicate the increase or decrease in Shannon diversity. For the following categorical variables: Cities are in reference to Barranquilla city, and “Sex (Male)” is in reference to “Sex (Female)”. Significant results are indicated by an asterisk (\* =  $P < 0.05$ , \*\* =  $P < 0.01$ , \*\*\* =  $P < 0.001$ , \*\*\*\* =  $P < 0.0001$ ).  $n = 361$ .

Metadata Category	Estimate	P-Value	Significance
Age (Years)	0.011	0.128	ns
City (Bogota)	0.494	0.000	****
City (Bucaramana)	0.029	0.843	ns
City (Cali)	-0.245	0.048	*
City (Medellin)	0.411	0.003	**
Carbohydrates (% Daily Value)	-0.075	0.017	*

Total Fat (% Daily Value)	-0.089	0.012	*
Animal Protein (% Daily Value)	-0.029	0.001	**
Sex (Male)	-0.221	0.106	ns

---