NMR Food (Short)

## Setup & Data

#| label: setup library(readxl) library(dplyr) library(tidyr) library(ggplot2)

# Path to your Excel (prefer a relative path like “data/CollectionNMRFood.xlsx”)

xlsx\_path <- “data/CollectionNMRFood.xlsx”

raw <- read\_excel(xlsx\_path, sheet = “datidef”) |> filter(ppm >= 0, ppm < 200)

# Matrix of samples only (drop ppm), rownames = ppm

mat <- as.data.frame(raw[,-1]) rownames(mat) <- raw$ppm

# Column-normalize (each sample sums to 1)

cs <- colSums(mat, na.rm = TRUE) norm <- sweep(mat, 2, cs, /) |> tibble::rownames\_to\_column(var = “ppm”) |> mutate(ppm = as.numeric(ppm))

#| label: classes classes <- cut( norm$ppm, breaks = c(-Inf, 45, 60, 90, 110, 145, 160, Inf), labels = c(“Alkyl C”, “Methoxyl N-Alkyl-C”, “O-alkyl-C”, “Di-O-alkyl C”, “H- C- sub. aromatic C”, “O- sub. aromatic C”, “Carbonyl C”), right = TRUE )

norm <- mutate(norm, MolecularClasses = classes)

# Sum by class (keep sample columns only)

summed <- norm |> group\_by(MolecularClasses) |> summarise(across(where(is.numeric) & !ppm, ~ sum(.x, na.rm = TRUE))) |> ungroup()

# Order samples by O-alkyl-C contribution (if present)

sample\_cols <- setdiff(names(summed), “MolecularClasses”) if (“O-alkyl-C” %in% summed$MolecularClasses) {
oidx <- which(summed$MolecularClasses == “O-alkyl-C”) ord <- order(-unlist(summed[oidx, sample\_cols])) summed <- select(summed, MolecularClasses, all\_of(sample\_cols[ord])) }

#| label: composition #| fig-width: 8 #| fig-height: 4 long <- pivot\_longer(summed, -MolecularClasses, names\_to = “sample”, values\_to = “value”)

ggplot(long, aes(sample, value, fill = MolecularClasses)) + geom\_bar(stat = “identity”, position = “fill”, color = “black”, linewidth = 0.1) + scale\_y\_continuous(labels = scales::percent) + labs(x = NULL, y = NULL) + theme\_classic() + theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust = 1))