AE 19: Increase in cost-burdened households in the United States

Suggested answers

APPLICATION EXERCISE ANSWERS

MODIFIED
April 10, 2025

```
library(tidyverse)
library(plotly)
library(scales)
library(colorspace)
library(ggrepel)

theme_set(theme_minimal())
```

We have already seen this semester that the cost of housing in the United States has been rising for several decades. A household is considered **cost-burdened** if they spend more than 30% of their income on housing costs.

In this application exercise we will explore trends in the percentage of cost-burdened rental households in the 10 largest metropolitan statistical areas (MSAs).[1] The relevant data can be found in `data/msa-renters-burden.csv`.

```
renter_burden <- read_csv(file = "data/msa-renters-burden.csv")
renter_burden
```

```
# A tibble: 110 × 4
    year geoid name                                      pct_burdened
   <dbl> <dbl> <chr>                                            <dbl>
 1  2013 12060 Atlanta-Sandy Springs-Roswell, GA                0.500
 2  2013 16980 Chicago-Naperville-Elgin, IL-IN                  0.493
 3  2013 19100 Dallas-Fort Worth-Arlington, TX                  0.453
 4  2013 26420 Houston-Pasadena-The Woodlands, TX               0.460
 5  2013 31080 Los Angeles-Long Beach-Anaheim, CA               0.561
 6  2013 33100 Miami-Fort Lauderdale-West Palm Beach, FL        0.595
 7  2013 35620 New York-Newark-Jersey City, NY-NJ               0.511
 8  2013 37980 Philadelphia-Camden-Wilmington, PA-NJ-DE-MD      0.503
 9  2013 38060 Phoenix-Mesa-Chandler, AZ                        0.477
```

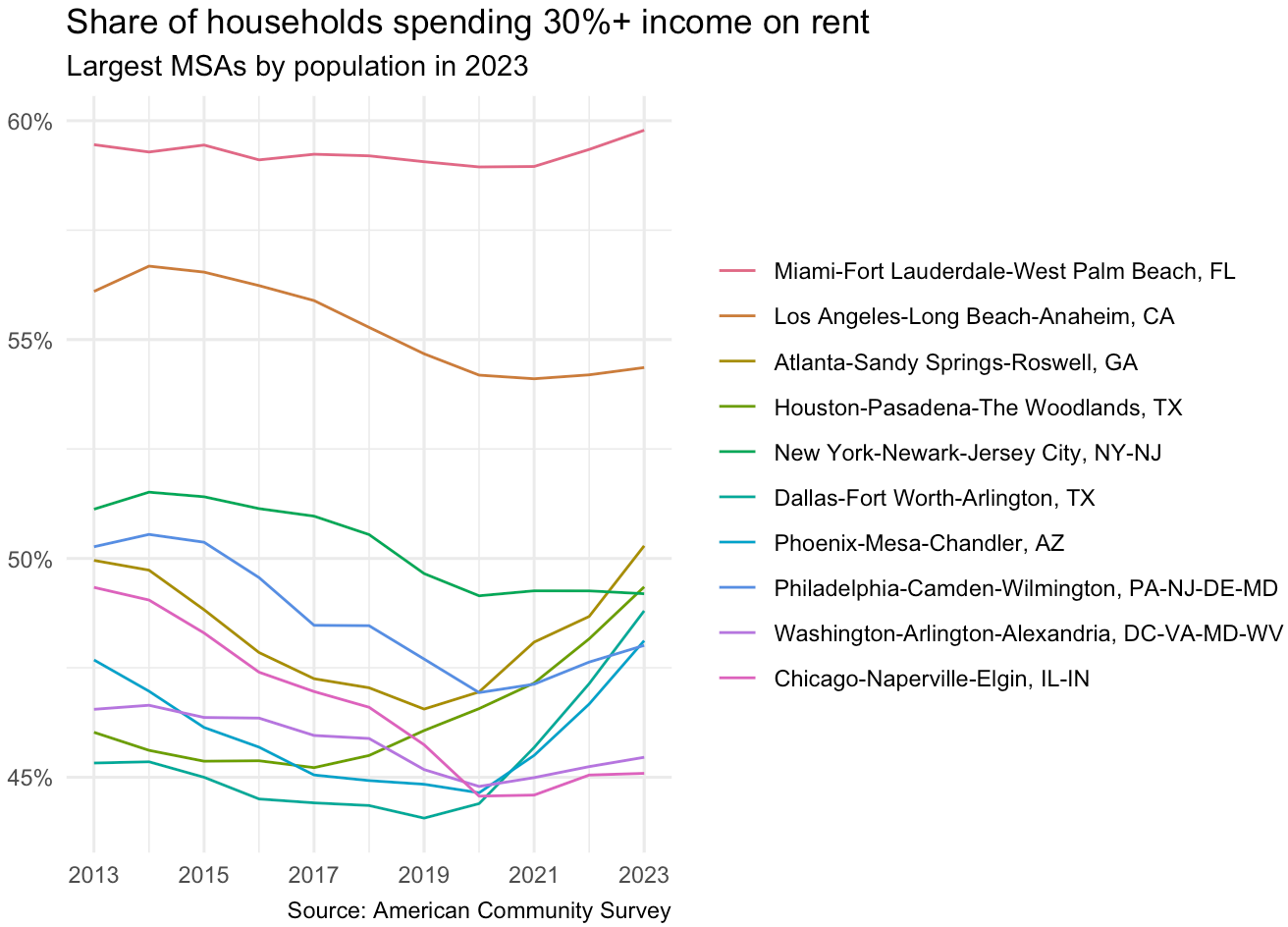
```
10  2013 47900 Washington-Arlington-Alexandria, DC-VA-MD-WV        0.466
# ℹ 100 more rows
```

`pct_burdened` reports the percentage of renter-occupied housing units that spend 30%+ of their household income on gross rent.[2]

Communicating trends with a static visualization

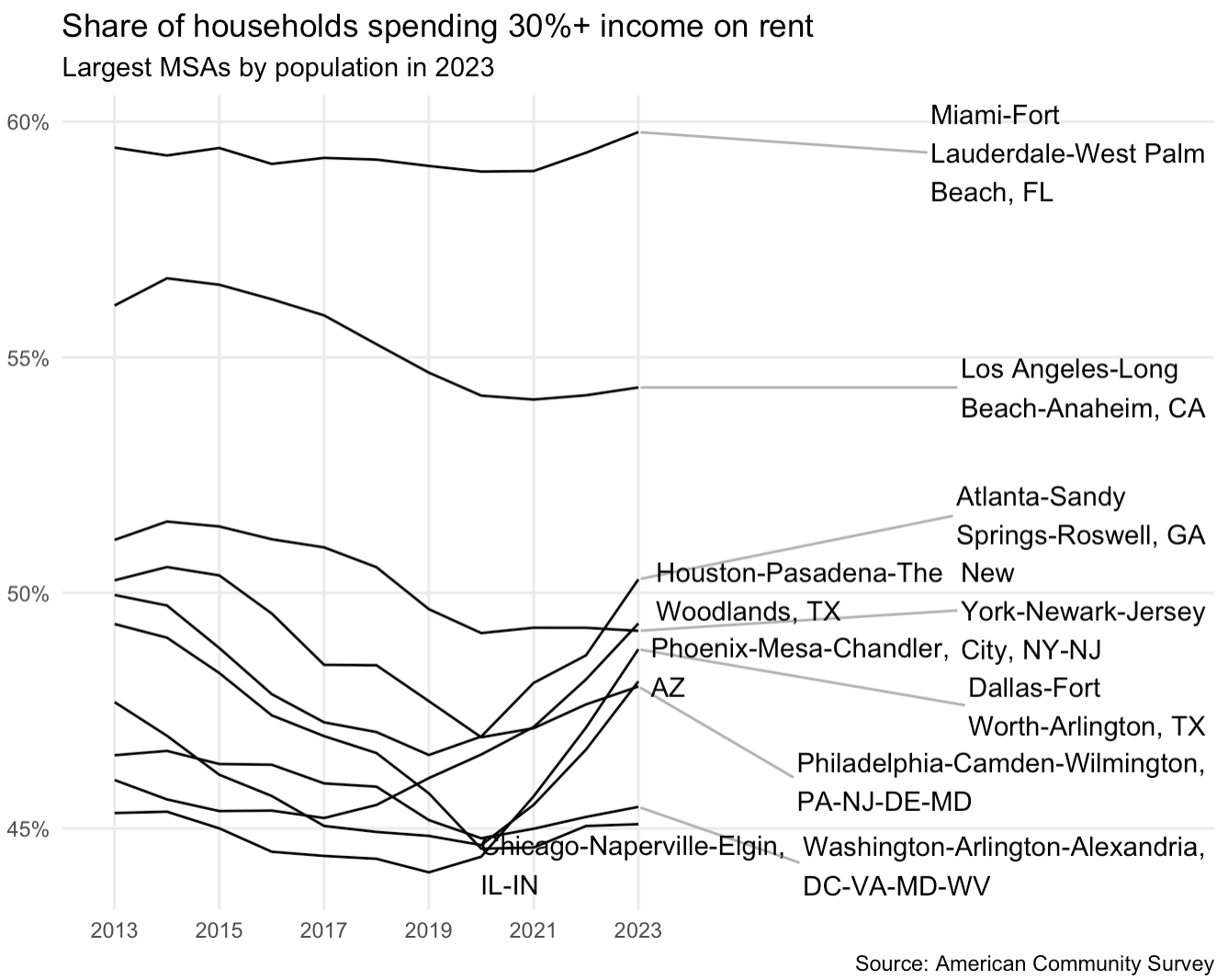
Your turn: While Americans face rising housing costs, the percentage of cost-burdened households has not increased uniformly across the country. Design and implement a static visualization to communicate the trends for these 10 MSAs. Ensure it can reasonably be used to identify trends specific to each MSA.

```
# use color to distinguish between MSAs
renter_burden |>
  # order the names by the most recent year for improved clarity in the legend
  mutate(name = fct_reorder2(.f = name, .x = year, .y = pct_burdened)) |>
  ggplot(mapping = aes(x = year, y = pct_burdened, color = name)) +
  geom_line() +
  # better x-axis breaks
  scale_x_continuous(breaks = seq(2013, 2023, by = 2)) +
  scale_y_continuous(labels = label_percent()) +
  scale_color_discrete_qualitative() +
  labs(
    x = NULL,
    y = NULL,
    color = NULL,
    title = "Share of households spending 30%+ income on rent",
    subtitle = "Largest MSAs by population in 2023",
    caption = "Source: American Community Survey"
  )
```




```
# set seed for reproducibility
set.seed(123)
ggplot(data = renter_burden, mapping = aes(x = year, y = pct_burdened, group = name)) +
  geom_line() +
  # just label the last year in the data
  geom_text_repel(
    data = renter_burden |>
      slice_max(order_by = year, n = 1, by = geoid),
    # wrap the character strings for space
    mapping = aes(label = str_wrap(name, width = 20)),
    # move labels over 10 units to the right before repelling them
    nudge_x = 10,
    # left alignment of text
    hjust = 0,
    segment.color = "grey70"
  ) +
  scale_x_continuous(breaks = seq(2013, 2023, by = 2)) +
  scale_y_continuous(labels = label_percent()) +
  labs(
    x = NULL,
```

```
    y = NULL,
    title = "Share of households spending 30%+ income on rent",
    subtitle = "Largest MSAs by population in 2023",
    caption = "Source: American Community Survey"
  ) +
  theme(panel.grid.minor = element_blank())
```



Communicating trends with an interactive visualization

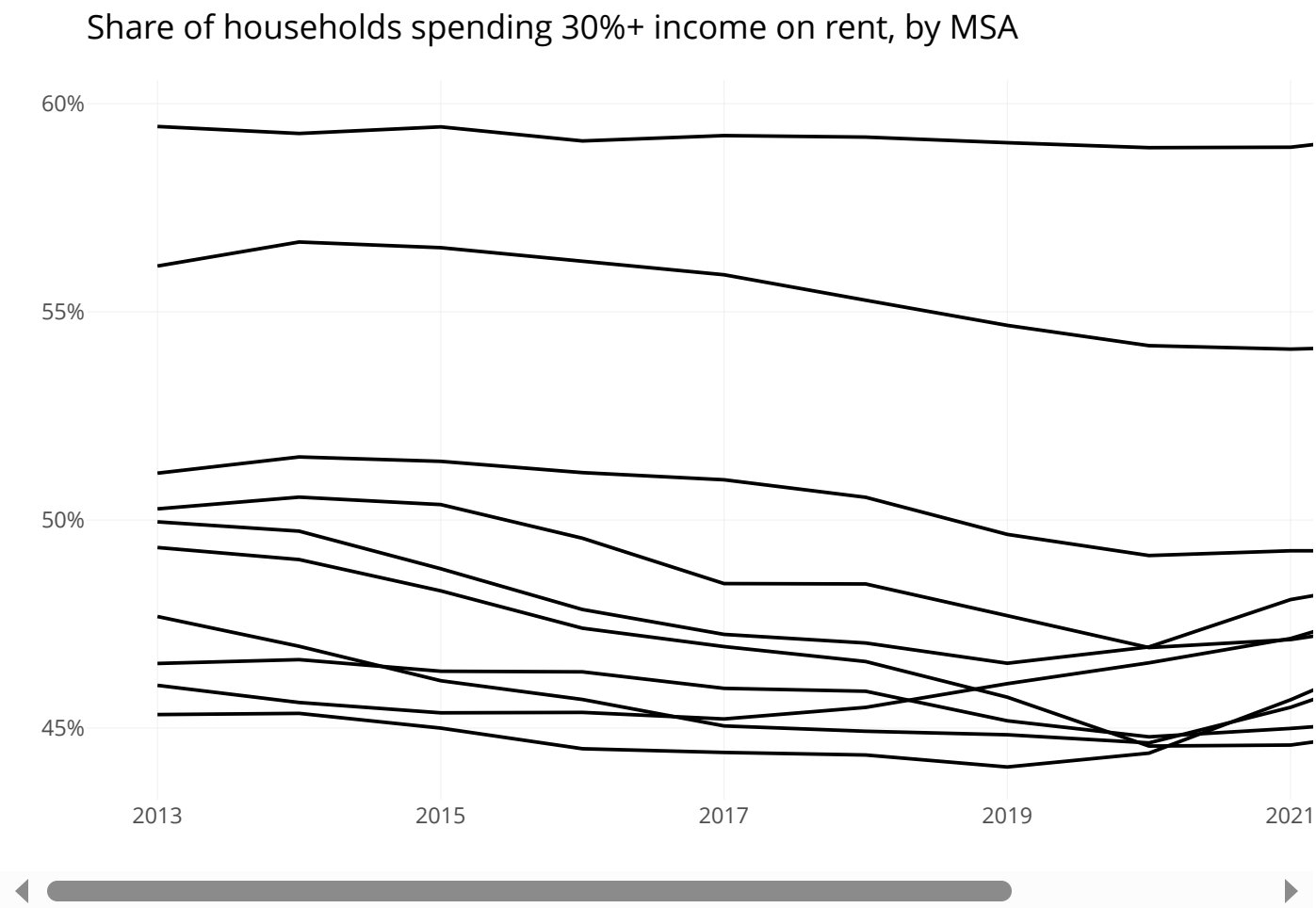
Your turn: Design and implement an interactive visualization to communicate the trends for these 10 MSAs. Ensure it can reasonably be used to identify trends specific to each MSA. Leverage interactive components to reduce clutter in the visualization and effectively utilize interactivity.

Suggestions include

- Customizing the tooltip to provide better-formatted information
- `highlight()` trend lines to draw attention to selected MSA
- Implement the plot purely using `plot_ly()`

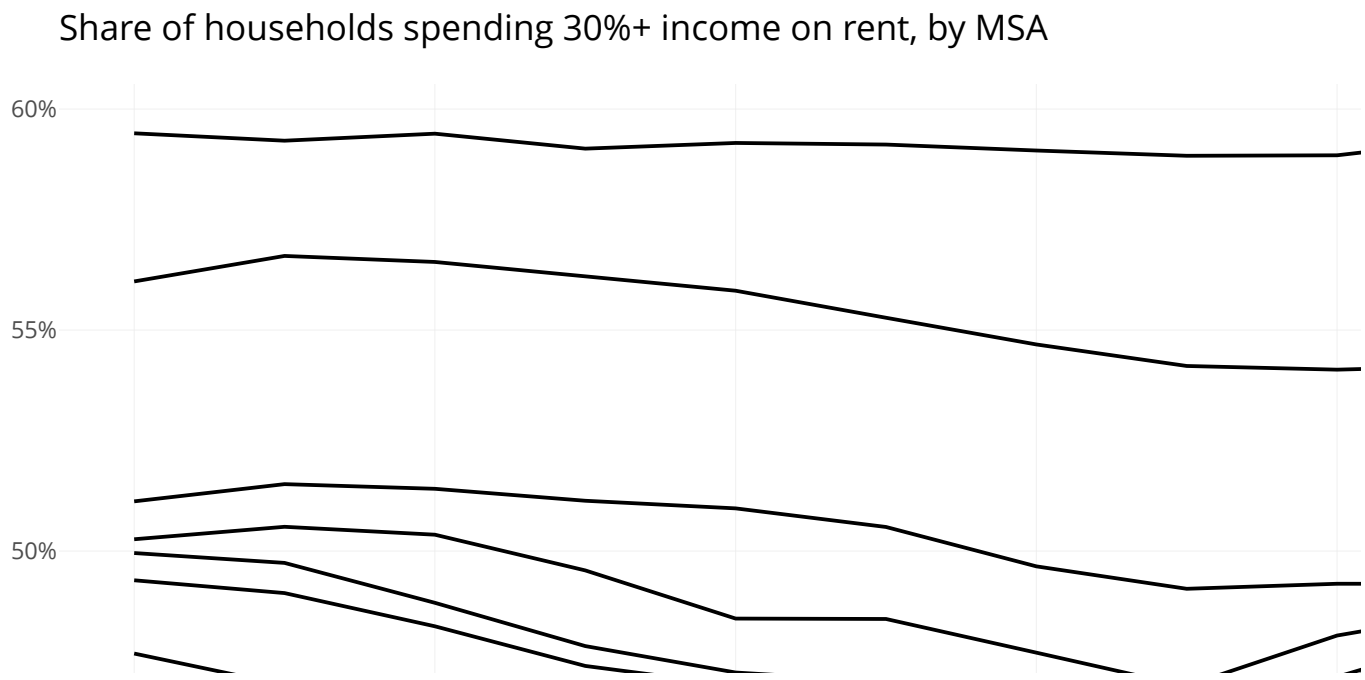
Using `ggplotly()` and customizing the tooltip

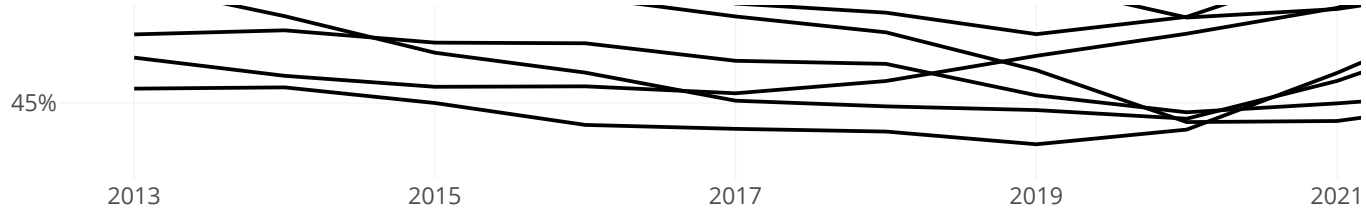
```
p <- renter_burden |>
  # create custom tooltip
  mutate(tooltip = str_glue("{name}<br>Year: {year}<br>Share of cost-burdened renters:
         {label_percent(accuracy = 1)(pct_burdened)}")) |>
  ggplot(mapping = aes(x = year, y = pct_burdened, group = name)) +
  geom_line(mapping = aes(text = tooltip)) +
  scale_x_continuous(breaks = seq(2013, 2023, by = 2)) +
  scale_y_continuous(labels = label_percent()) +
  labs(
    x = NULL,
    y = NULL,
    title = "Share of households spending 30%+ income on rent, by MSA"
  )

# render with plotly
ggplotly(p, tooltip = "text")
```



Highlight a specific line

```
{
  # create tooltip
  renter_burden |>
    mutate(tooltip = str_glue("{name}<br>Year: {year}<br>Share of cost-burdened renters:
          {label_percent(accuracy = 1)(pct_burdened)}")) |>
    # create a SharedData object for use in the ggplot() below and group by name
    highlight_key(~name) |>
    # create the static plot
    ggplot(mapping = aes(x = year, y = pct_burdened, group = name)) +
    geom_line(mapping = aes(text = tooltip)) +
    scale_x_continuous(breaks = seq(2013, 2023, by = 2)) +
    scale_y_continuous(labels = label_percent()) +
    labs(
      x = NULL,
      y = NULL,
      title = "Share of households spending 30%+ income on rent, by MSA"
    )
} |>
  # convert to interactive plot
  ggplotly(tooltip = "text") |>
  # activate highlight on hover
  highlight(
    # events triggering highlight on and off
    on = "plotly_hover",
    off = "plotly_doubleclick",
    # highlight color
    color = "orange"
  )
```

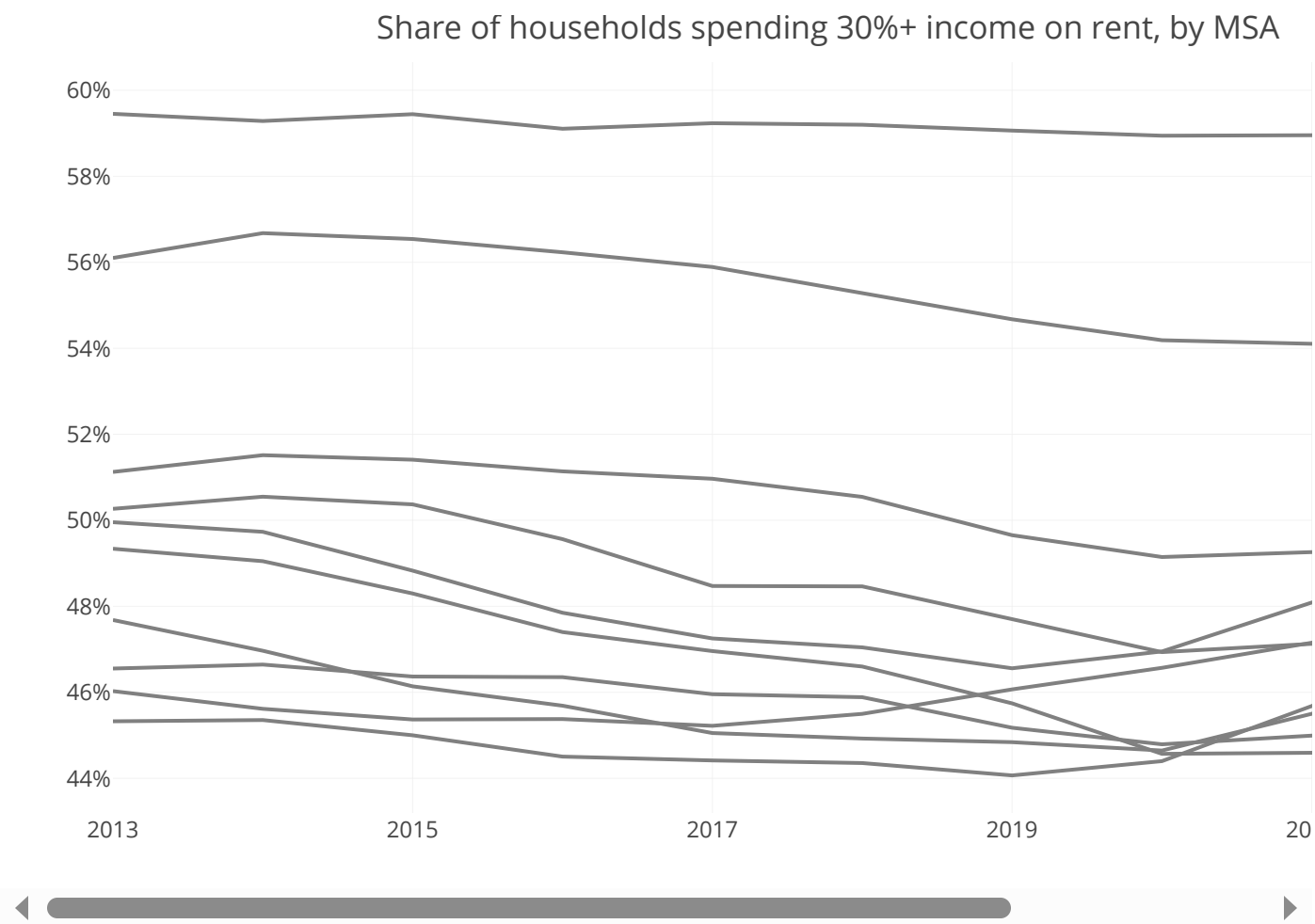




Implement using `plot_ly()`

```
# Load required libraries
library(plotly)
library(dplyr)
library(stringr)
library(scales)
library(crosstalk)  # for highlight_key()

# Build interactive line plot with highlighting using plotly
renter_burden |>
  # Create a custom tooltip for each data point
  # Tooltip includes metro area name, year, and the percent of cost-burdened renters
  mutate(tooltip = str_glue(
    "{name}<br>Year: {year}<br>Share of cost-burdened renters: {label_percent(accuracy = 1)
          (pct_burdened)}"
  )) |>

  # Register 'name' as the group key for interactivity
  # This allows plotly to know which line each point belongs to
  highlight_key(~name) |>

  # Create the interactive plot
  plot_ly(
    x = ~year,                  # x-axis: year
    y = ~pct_burdened,          # y-axis: percent cost-burdened
    text = ~tooltip,            # use custom tooltip text
    hoverinfo = "text",         # only show the text we defined, not default info
    type = "scatter",           # create a scatterplot (with lines below)
    mode = "lines",             # connect points with lines
    split = ~name,              # create a separate trace for each metro area (MSA)
    line = list(color = "gray"), # set all lines to gray by default
    showlegend = FALSE          # don't show legend for each metro area
  ) |>

  # Customize plot layout
```

```
  layout(
    title = "Share of households spending 30%+ income on rent, by MSA",
    xaxis = list(
      title = NA,
      tickvals = seq(2013, 2023, by = 2)  # show ticks every 2 years
    ),
    yaxis = list(
      title = NA,
      tickformat = ".0%",  # format y-axis as percentages
      rangemode = "normal" # ensure full range is shown
    ),
    hovermode = "closest",   # only show tooltip for the closest point being hovered
    showlegend = FALSE       # make sure legend stays hidden
  ) |>

  # Enable interactive highlighting
  highlight(
    on = "plotly_hover",        # highlight a line when user hovers over it
    off = "plotly_doubleclick", # reset highlights when user double-clicks
    color = "orange"            # highlight color for selected line
  )
```



Session information

Footnotes

1. Based on population as of 2023. ↩
2. Specifically Table B25070 from the American Community Survey. ↩