# south_end_complete

February 15, 2022

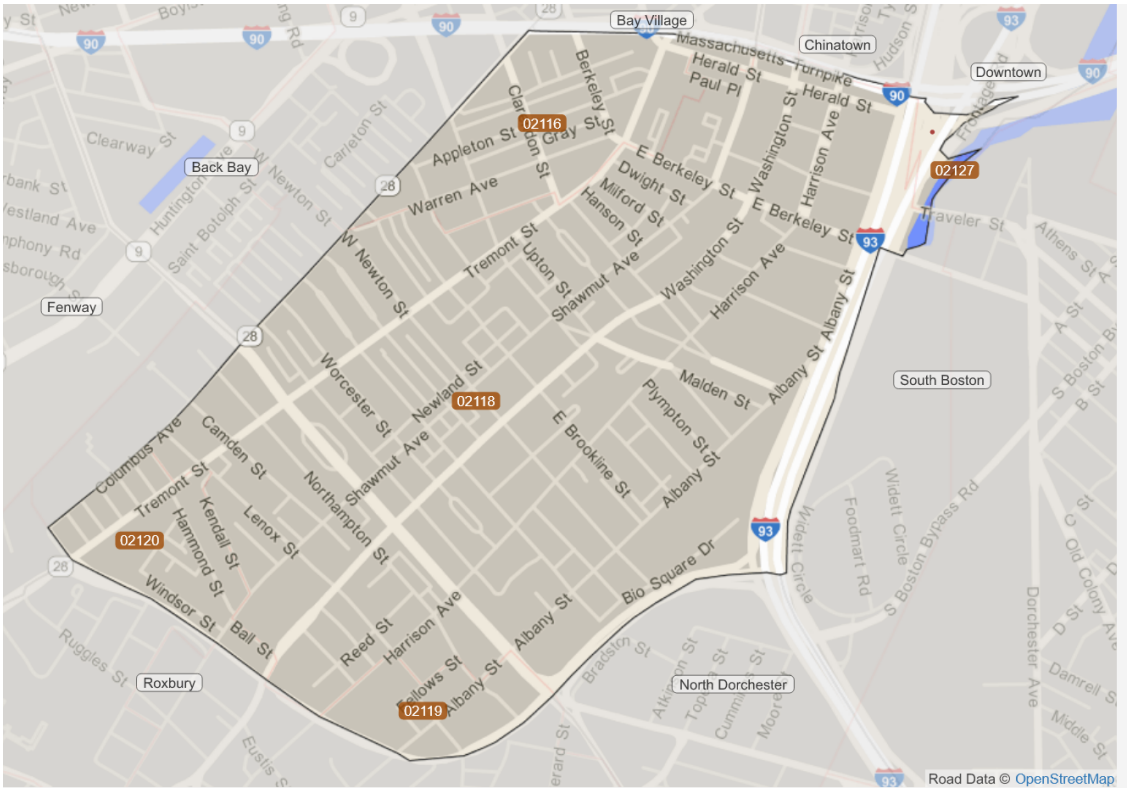
## 1 Keyboard Shortcuts

Cell organization mode * `enter` to begin editing a cell * `up` and `down` arrows to move between cells * `D,D` to delete a cell * `M` to change to Markdown * `Y` to change to code * `C` to copy a cell * `V` to paste the cell below * `B` insert cell below * `A` insert cell above

Editing mode * `esc` to exit editing the cell * `ctrl+enter` to execute a cell * `alt+enter` to execute a cell and create a new one below * `shift+enter` to execute a cell and move to the next one * `ctrl+/` comment the selected line(s)

## 2 Background

The census tract for the South End begins just on the SE side of Columbus Avenue, as shown in this map:



Let's explore the demographics of the South End.

## 3 Setup

[1]:
```
import pandas as pd
import altair as alt
from altair import datum

alt.renderers.enable('jupyterlab', embed_options={'renderer': 'svg'})

# Avoids writing all the data to the notebook or disk.
# See https://altair-viz.github.io/user_guide/faq.html#local-data-server
# Note that this may not work on some cloud-based Jupyter notebook services.
# alt.data_transformers.enable('data_server')
```

[1]:
```
RendererRegistry.enable('jupyterlab')
```

## 4 Loading Data

[2]:
```
df = pd.read_csv('south_end.csv')
df
```

[2]:

| | Category | Subcategory | Decade | Count | Percent |
|---|---|---|---|---|---|
| 0 | Population | NaN | 1950 | 49753 | NaN |
| 1 | Age | 0-9 years | 1950 | 5870 | 0.12 |
| 2 | Age | 10-19 years | 1950 | 4387 | 0.09 |
| 3 | Age | 20-34 years | 1950 | 11947 | 0.24 |
| 4 | Age | 35-54 years | 1950 | 14374 | 0.29 |
| .. | … | … | … | … | … |
| 149 | Labor Force (age 16+) | Male | 2010 | - | NaN |
| 150 | Labor Force (age 16+) | Female | 2010 | - | NaN |
| 151 | Housing Tenure | Occupied Housing Units | 2010 | 15629 | NaN |
| 152 | Housing Tenure | Owner-occupied | 2010 | 5702 | 0.36 |
| 153 | Housing Tenure | Renter-occupied | 2010 | 9927 | 0.64 |

[154 rows x 5 columns]

[3]:
```
pd.set_option('display.max_rows', None) # 60 is the default
df[['Category', 'Subcategory']]
```

[3]:

| | Category | Subcategory |
|---|---|---|
| 0 | Population | NaN |
| 1 | Age | 0-9 years |
| 2 | Age | 10-19 years |
| 3 | Age | 20-34 years |
| 4 | Age | 35-54 years |

| | | | |
|---|---|---|---|
| 5 | | Age | 55-64 years |
| 6 | | Age | 65 years and over |
| 7 | Educational Attainment (age 25+) | | less than High School |
| 8 | Educational Attainment (age 25+) | | High School or GED |
| 9 | Educational Attainment (age 25+) | | Some College or Associate's Degree |
| 10 | Educational Attainment (age 25+) | | Bachelor's Degree or Higher |
| 11 | | Nativity | Foreign Born |
| 12 | | Race/ Ethnicity | White |
| 13 | | Race/ Ethnicity | Black/ African American |
| 14 | | Race/ Ethnicity | Hispanic |
| 15 | | Race/ Ethnicity | Asian/PI |
| 16 | | Race/ Ethnicity | Other |
| 17 | | Labor Force (age 16+) | Male |
| 18 | | Labor Force (age 16+) | Female |
| 19 | | Housing Tenure | Occupied Housing Units |
| 20 | | Housing Tenure | Owner-occupied |
| 21 | | Housing Tenure | Renter-occupied |
| 22 | | Population | NaN |
| 23 | | Age | 0-9 years |
| 24 | | Age | 10-19 years |
| 25 | | Age | 20-34 years |
| 26 | | Age | 35-54 years |
| 27 | | Age | 55-64 years |
| 28 | | Age | 65 years and over |
| 29 | Educational Attainment (age 25+) | | less than High School |
| 30 | Educational Attainment (age 25+) | | High School or GED |
| 31 | Educational Attainment (age 25+) | | Some College or Associate's Degree |
| 32 | Educational Attainment (age 25+) | | Bachelor's Degree or Higher |
| 33 | | Nativity | Foreign Born |
| 34 | | Race/ Ethnicity | White |
| 35 | | Race/ Ethnicity | Black/ African American |
| 36 | | Race/ Ethnicity | Hispanic |
| 37 | | Race/ Ethnicity | Asian/PI |
| 38 | | Race/ Ethnicity | Other |
| 39 | | Labor Force (age 16+) | Male |
| 40 | | Labor Force (age 16+) | Female |
| 41 | | Housing Tenure | Occupied Housing Units |
| 42 | | Housing Tenure | Owner-occupied |
| 43 | | Housing Tenure | Renter-occupied |
| 44 | | Population | NaN |
| 45 | | Age | 0-9 years |
| 46 | | Age | 10-19 years |
| 47 | | Age | 20-34 years |
| 48 | | Age | 35-54 years |
| 49 | | Age | 55-64 years |
| 50 | | Age | 65 years and over |
| 51 | Educational Attainment (age 25+) | | less than High School |

| | | |
|---|---|---|
| 52 | Educational Attainment (age 25+) | High School or GED |
| 53 | Educational Attainment (age 25+) | Some College or Associate's Degree |
| 54 | Educational Attainment (age 25+) | Bachelor's Degree or Higher |
| 55 | Nativity | Foreign Born |
| 56 | Race/ Ethnicity | White |
| 57 | Race/ Ethnicity | Black/ African American |
| 58 | Race/ Ethnicity | Hispanic |
| 59 | Race/ Ethnicity | Asian/PI |
| 60 | Race/ Ethnicity | Other |
| 61 | Labor Force (age 16+) | Male |
| 62 | Labor Force (age 16+) | Female |
| 63 | Housing Tenure | Occupied Housing Units |
| 64 | Housing Tenure | Owner-occupied |
| 65 | Housing Tenure | Renter-occupied |
| 66 | Population | NaN |
| 67 | Age | 0-9 years |
| 68 | Age | 10-19 years |
| 69 | Age | 20-34 years |
| 70 | Age | 35-54 years |
| 71 | Age | 55-64 years |
| 72 | Age | 65 years and over |
| 73 | Educational Attainment (age 25+) | less than High School |
| 74 | Educational Attainment (age 25+) | High School or GED |
| 75 | Educational Attainment (age 25+) | Some College or Associate's Degree |
| 76 | Educational Attainment (age 25+) | Bachelor's Degree or Higher |
| 77 | Nativity | Foreign Born |
| 78 | Race/ Ethnicity | White |
| 79 | Race/ Ethnicity | Black/ African American |
| 80 | Race/ Ethnicity | Hispanic |
| 81 | Race/ Ethnicity | Asian/PI |
| 82 | Race/ Ethnicity | Other |
| 83 | Labor Force (age 16+) | Male |
| 84 | Labor Force (age 16+) | Female |
| 85 | Housing Tenure | Occupied Housing Units |
| 86 | Housing Tenure | Owner-occupied |
| 87 | Housing Tenure | Renter-occupied |
| 88 | Population | NaN |
| 89 | Age | 0-9 years |
| 90 | Age | 10-19 years |
| 91 | Age | 20-34 years |
| 92 | Age | 35-54 years |
| 93 | Age | 55-64 years |
| 94 | Age | 65 years and over |
| 95 | Educational Attainment (age 25+) | less than High School |
| 96 | Educational Attainment (age 25+) | High School or GED |
| 97 | Educational Attainment (age 25+) | Some College or Associate's Degree |
| 98 | Educational Attainment (age 25+) | Bachelor's Degree or Higher |

| | | |
|---|---|---|
| 99 | Nativity | Foreign Born |
| 100 | Race/ Ethnicity | White |
| 101 | Race/ Ethnicity | Black/ African American |
| 102 | Race/ Ethnicity | Hispanic |
| 103 | Race/ Ethnicity | Asian/PI |
| 104 | Race/ Ethnicity | Other |
| 105 | Labor Force (age 16+) | Male |
| 106 | Labor Force (age 16+) | Female |
| 107 | Housing Tenure | Occupied Housing Units |
| 108 | Housing Tenure | Owner-occupied |
| 109 | Housing Tenure | Renter-occupied |
| 110 | Population | NaN |
| 111 | Age | 0-9 years |
| 112 | Age | 10-19 years |
| 113 | Age | 20-34 years |
| 114 | Age | 35-54 years |
| 115 | Age | 55-64 years |
| 116 | Age | 65 years and over |
| 117 | Educational Attainment (age 25+) | less than High School |
| 118 | Educational Attainment (age 25+) | High School or GED |
| 119 | Educational Attainment (age 25+) | Some College or Associate's Degree |
| 120 | Educational Attainment (age 25+) | Bachelor's Degree or Higher |
| 121 | Nativity | Foreign Born |
| 122 | Race/ Ethnicity | White |
| 123 | Race/ Ethnicity | Black/ African American |
| 124 | Race/ Ethnicity | Hispanic |
| 125 | Race/ Ethnicity | Asian/PI |
| 126 | Race/ Ethnicity | Other |
| 127 | Labor Force (age 16+) | Male |
| 128 | Labor Force (age 16+) | Female |
| 129 | Housing Tenure | Occupied Housing Units |
| 130 | Housing Tenure | Owner-occupied |
| 131 | Housing Tenure | Renter-occupied |
| 132 | Population | NaN |
| 133 | Age | 0-9 years |
| 134 | Age | 10-19 years |
| 135 | Age | 20-34 years |
| 136 | Age | 35-54 years |
| 137 | Age | 55-64 years |
| 138 | Age | 65 years and over |
| 139 | Educational Attainment (age 25+) | less than High School |
| 140 | Educational Attainment (age 25+) | High School or GED |
| 141 | Educational Attainment (age 25+) | Some College or Associate's Degree |
| 142 | Educational Attainment (age 25+) | Bachelor's Degree or Higher |
| 143 | Nativity | Foreign Born |
| 144 | Race/ Ethnicity | White |
| 145 | Race/ Ethnicity | Black/ African American |

| | | |
|---|---|---|
| 146 | Race/ Ethnicity | Hispanic |
| 147 | Race/ Ethnicity | Asian/PI |
| 148 | Race/ Ethnicity | Other |
| 149 | Labor Force (age 16+) | Male |
| 150 | Labor Force (age 16+) | Female |
| 151 | Housing Tenure | Occupied Housing Units |
| 152 | Housing Tenure | Owner-occupied |
| 153 | Housing Tenure | Renter-occupied |

[4]:
```
df[['Category', 'Subcategory']].drop_duplicates()
```

[4]:

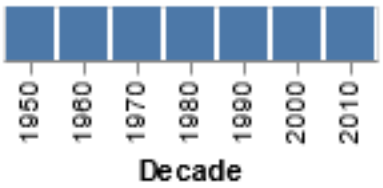
| | Category | Subcategory |
|---|---|---|
| 0 | Population | NaN |
| 1 | Age | 0-9 years |
| 2 | Age | 10-19 years |
| 3 | Age | 20-34 years |
| 4 | Age | 35-54 years |
| 5 | Age | 55-64 years |
| 6 | Age | 65 years and over |
| 7 | Educational Attainment (age 25+) | less than High School |
| 8 | Educational Attainment (age 25+) | High School or GED |
| 9 | Educational Attainment (age 25+) | Some College or Associate's Degree |
| 10 | Educational Attainment (age 25+) | Bachelor's Degree or Higher |
| 11 | Nativity | Foreign Born |
| 12 | Race/ Ethnicity | White |
| 13 | Race/ Ethnicity | Black/ African American |
| 14 | Race/ Ethnicity | Hispanic |
| 15 | Race/ Ethnicity | Asian/PI |
| 16 | Race/ Ethnicity | Other |
| 17 | Labor Force (age 16+) | Male |
| 18 | Labor Force (age 16+) | Female |
| 19 | Housing Tenure | Occupied Housing Units |
| 20 | Housing Tenure | Owner-occupied |
| 21 | Housing Tenure | Renter-occupied |

# 5 Creating Visualizations

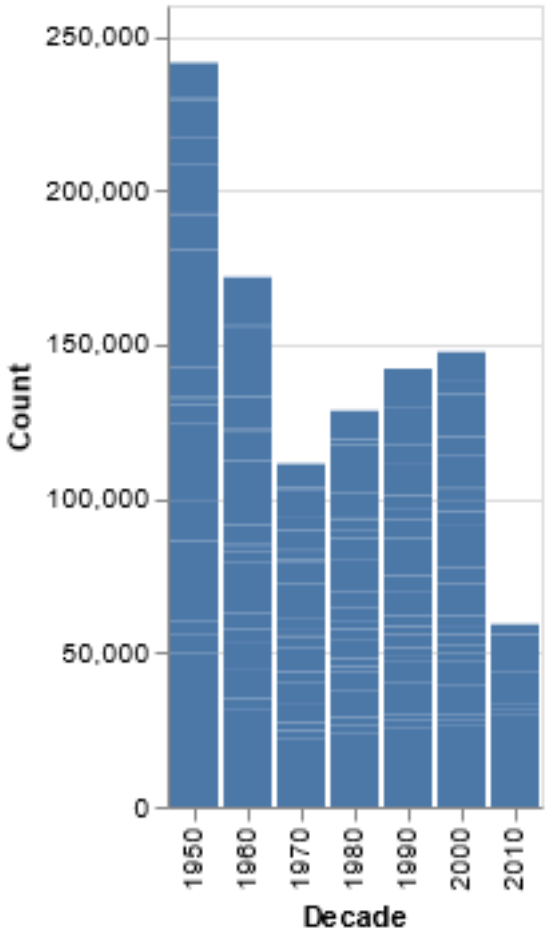
## 5.1 Age Distribution by Decade

[5]:
```
alt.Chart(df).mark_bar().encode(
    x='Decade:O'
)
```

[5]:



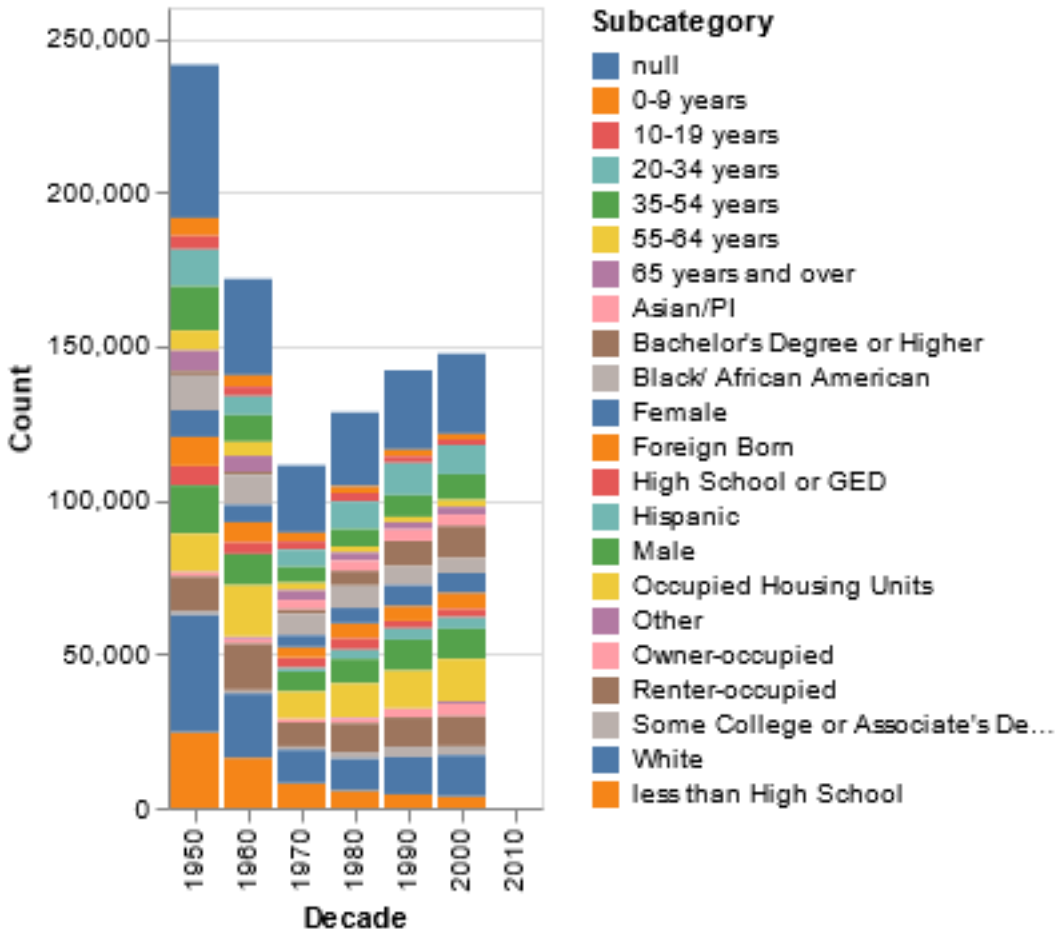

[6]:
```
alt.Chart(df).mark_bar().encode(
    x = 'Decade:O',
    y = 'Count:Q'
)
```

[6]:



[7]:
```
alt.Chart(df).mark_bar().encode(
    x = 'Decade:O',
    y = 'Count:Q',
    color = 'Subcategory:N'
```

```
)
```

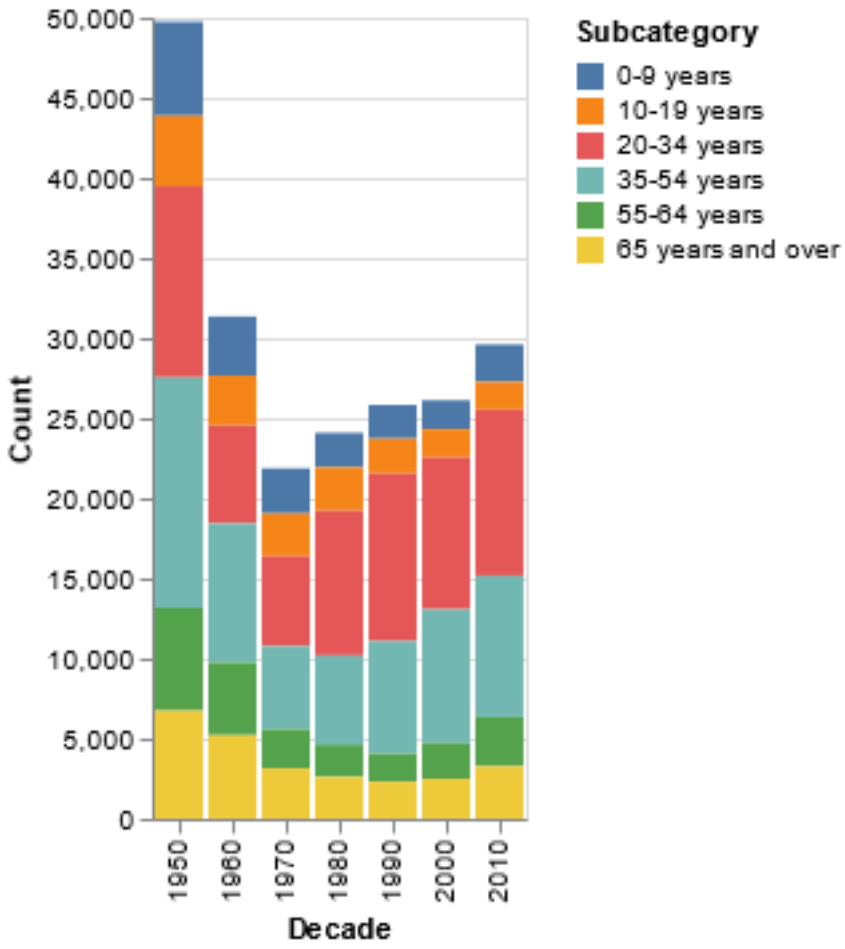
[7]:




[8]:

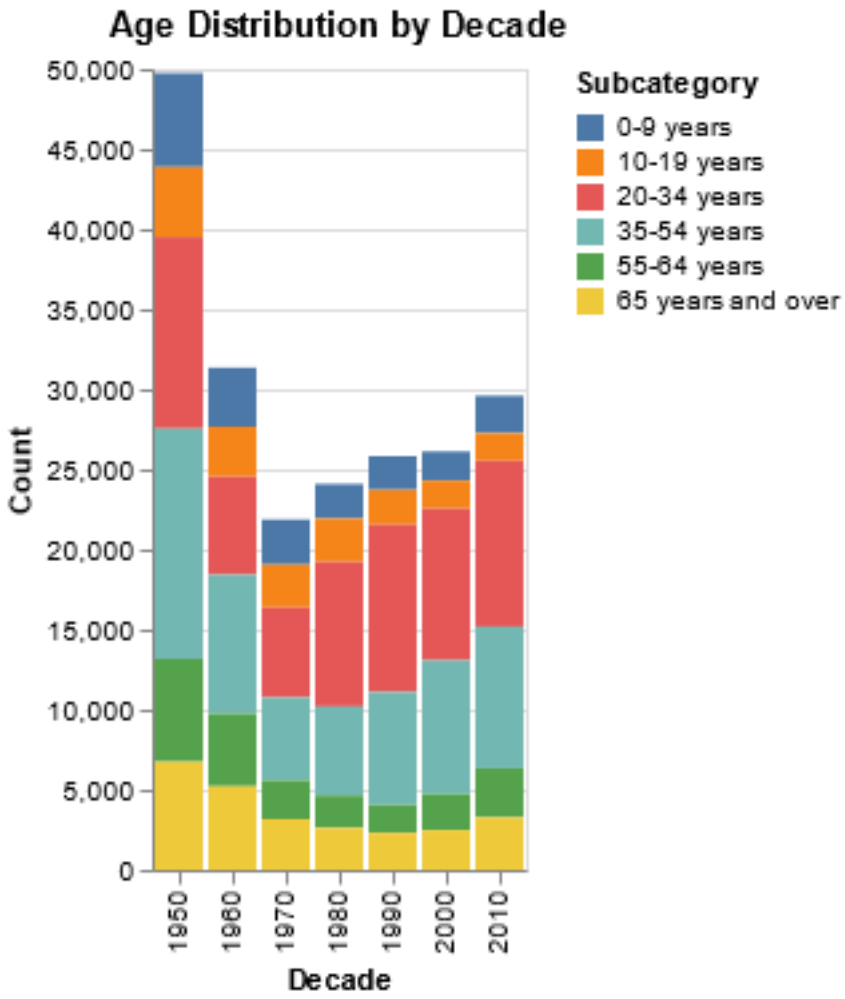
```
alt.Chart(df).mark_bar().encode(
    x = 'Decade:O',
    y = 'Count:Q',
    color = 'Subcategory:N'
).transform_filter(
    datum['Category'] == 'Age'
)
```

[8]:



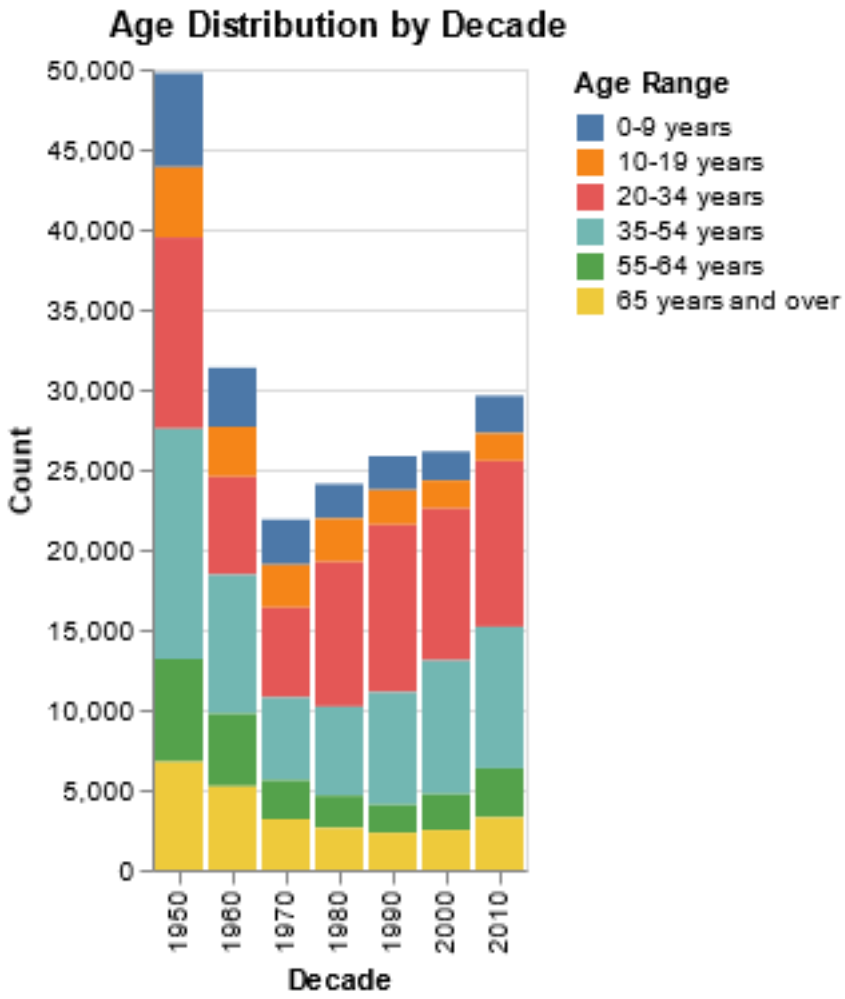

[9]:
```
alt.Chart(df).mark_bar().encode(
    x = 'Decade:O',
    y = 'Count:Q',
    color = 'Subcategory:N'
).transform_filter(
    datum['Category'] == 'Age'
).properties(
    title = 'Age Distribution by Decade'
)
```

[9]:



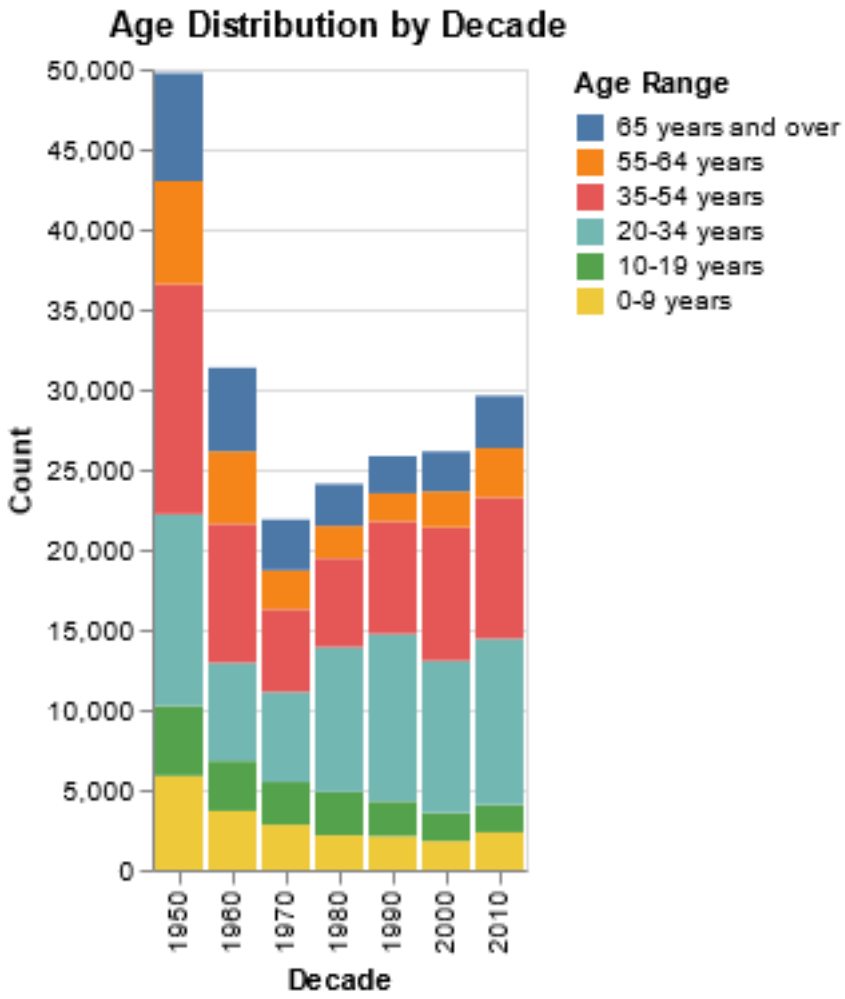

[10]:
```
alt.Chart(df).mark_bar().encode(
    x = 'Decade:O',
    y = 'Count:Q',
    color = alt.Color('Subcategory:N', title = 'Age Range')
).transform_filter(
    datum['Category'] == 'Age'
).properties(
    title = 'Age Distribution by Decade'
)
```

[10]:



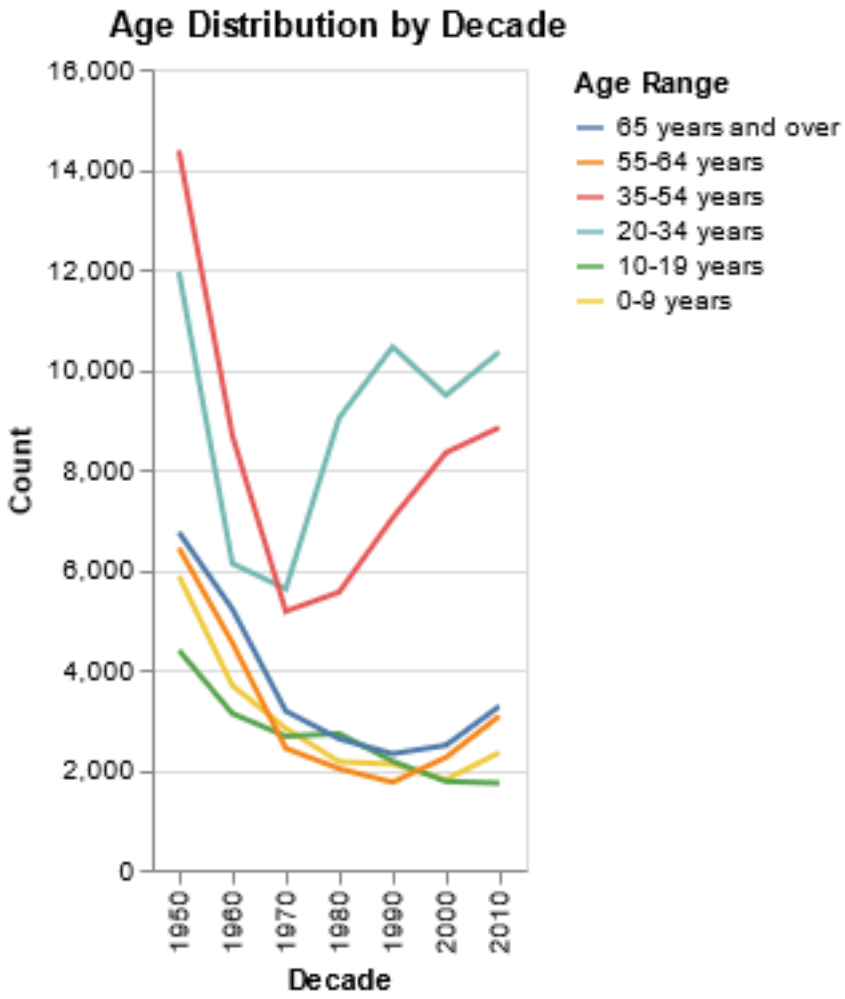

[11]:
```
alt.Chart(df).mark_bar().encode(
    x = 'Decade:O',
    y = 'Count:Q',
    color = alt.Color('Subcategory:N', title = 'Age Range', sort =␣
 ↪'descending'),
    order = alt.Order(
        'Subcategory:N',
        sort = 'ascending'
    )
).transform_filter(
    datum['Category'] == 'Age'
).properties(
    title = 'Age Distribution by Decade'
)
```

[11]:



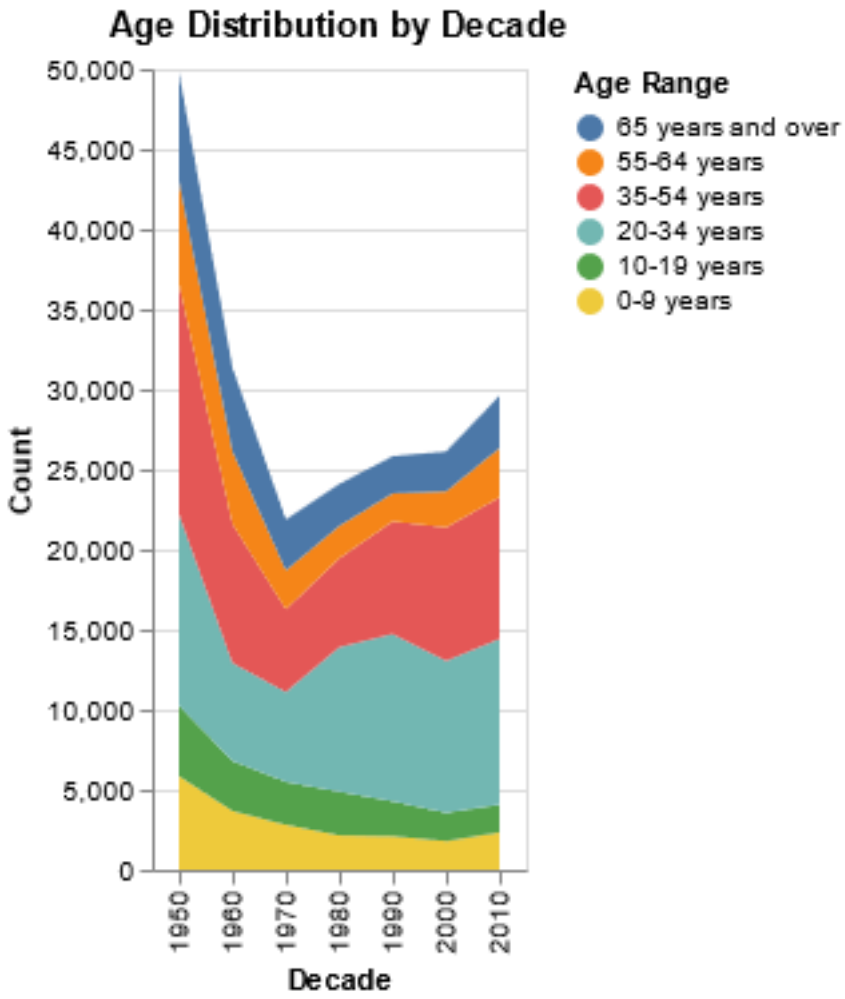

[12]:
```
alt.Chart(df).mark_line().encode(
    x = 'Decade:O',
    y = 'Count:Q',
    color = alt.Color('Subcategory:N', title = 'Age Range', sort =␣
 ↪'descending'),
    order = alt.Order(
        'Subcategory:N',
        sort = 'ascending'
    )
).transform_filter(
    datum['Category'] == 'Age'
).properties(
    title = 'Age Distribution by Decade'
)
```

[12]:




[13]:
```
alt.Chart(df).mark_area().encode(
    x = 'Decade:O',
    y = 'Count:Q',
    color = alt.Color('Subcategory:N', title = 'Age Range', sort =␣
 ↪'descending'),
    order = alt.Order(
        'Subcategory:N',
        sort = 'ascending'
    )
).transform_filter(
    datum['Category'] == 'Age'
).properties(
    title = 'Age Distribution by Decade'
)
```

[13]:



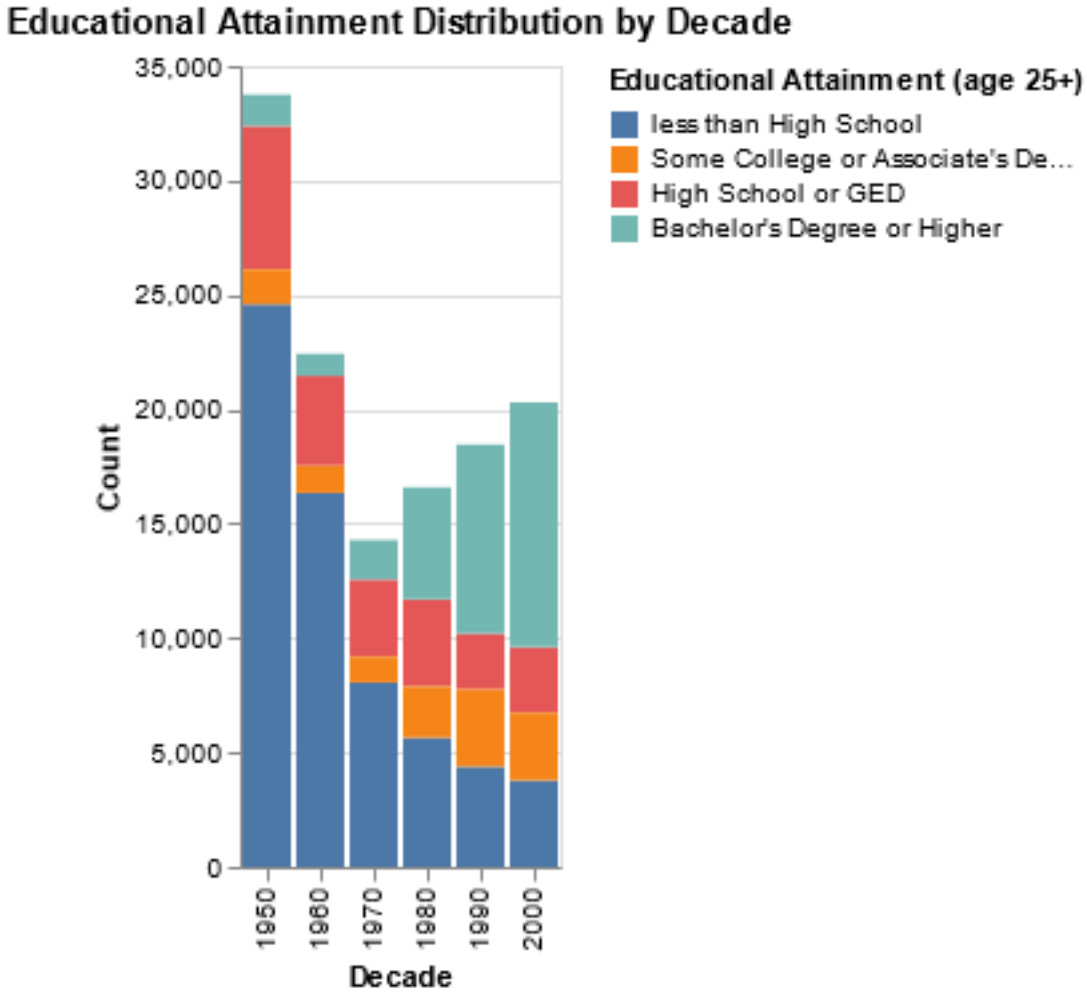

### 5.2 Educational Attainment by Decade

[14]:
```
eduField = 'Educational Attainment (age 25+)'
```

[15]:
```
alt.Chart(df).mark_bar().encode(
    x = 'Decade:O',
    y = 'Count:Q',
    color = alt.Color('Subcategory:N', title = eduField, sort = 'descending'),
    order = alt.Order(
        'Subcategory:N',
        sort = 'descending'
    )
).transform_filter(
    datum['Category'] == eduField
).properties(
    title = 'Educational Attainment Distribution by Decade'
```

```
)
```

[15]:

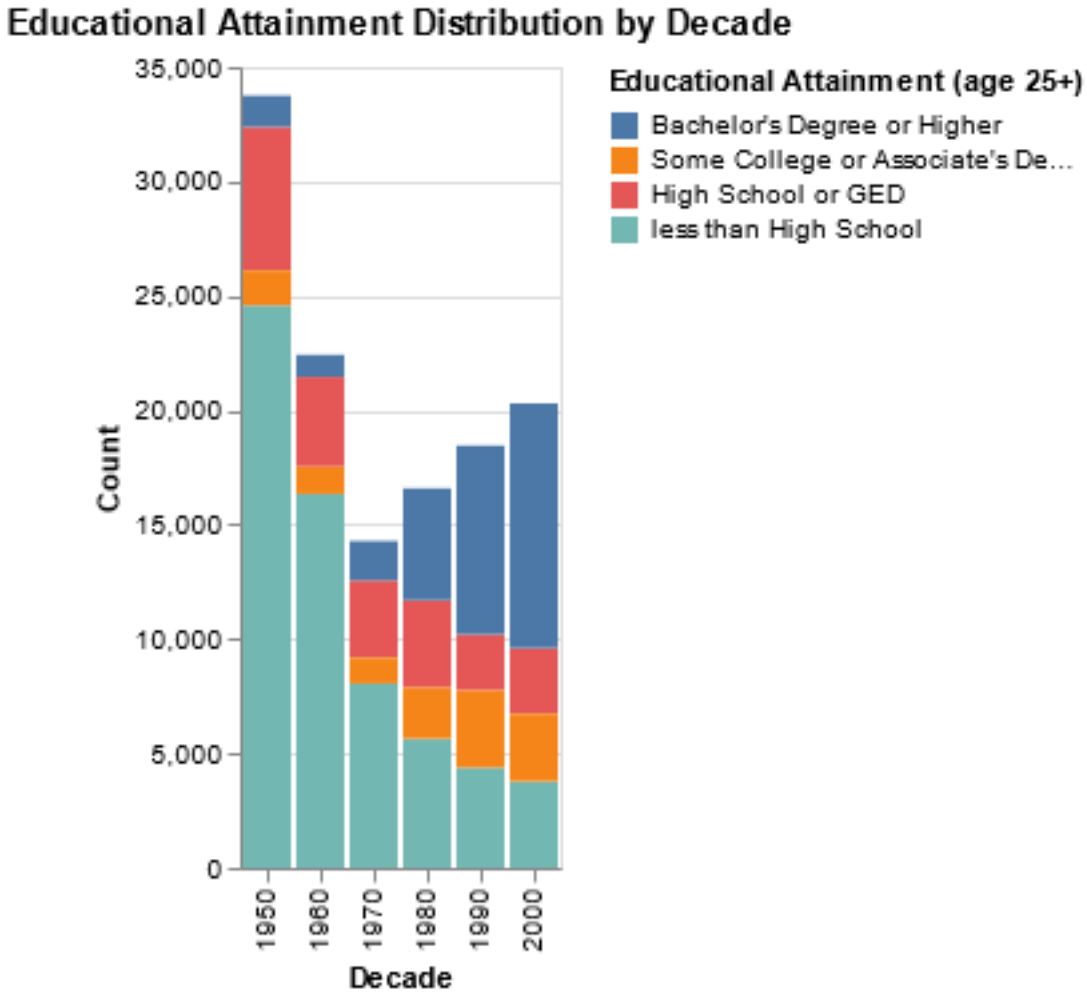



[16]:
```
eduSortOrder = [
        "Bachelor's Degree or Higher",
        "Some College or Associate's Degree",
        'High School or GED',
        'less than High School'
        ]
```

[17]:
```
alt.Chart(df).mark_bar().encode(
    x = 'Decade:O',
    y = 'Count:Q',
    color = alt.Color('Subcategory:N', title = eduField, sort = eduSortOrder),
).transform_filter(
    datum['Category'] == eduField
).properties(
```

```
    title = 'Educational Attainment Distribution by Decade'
)
```

[17]:

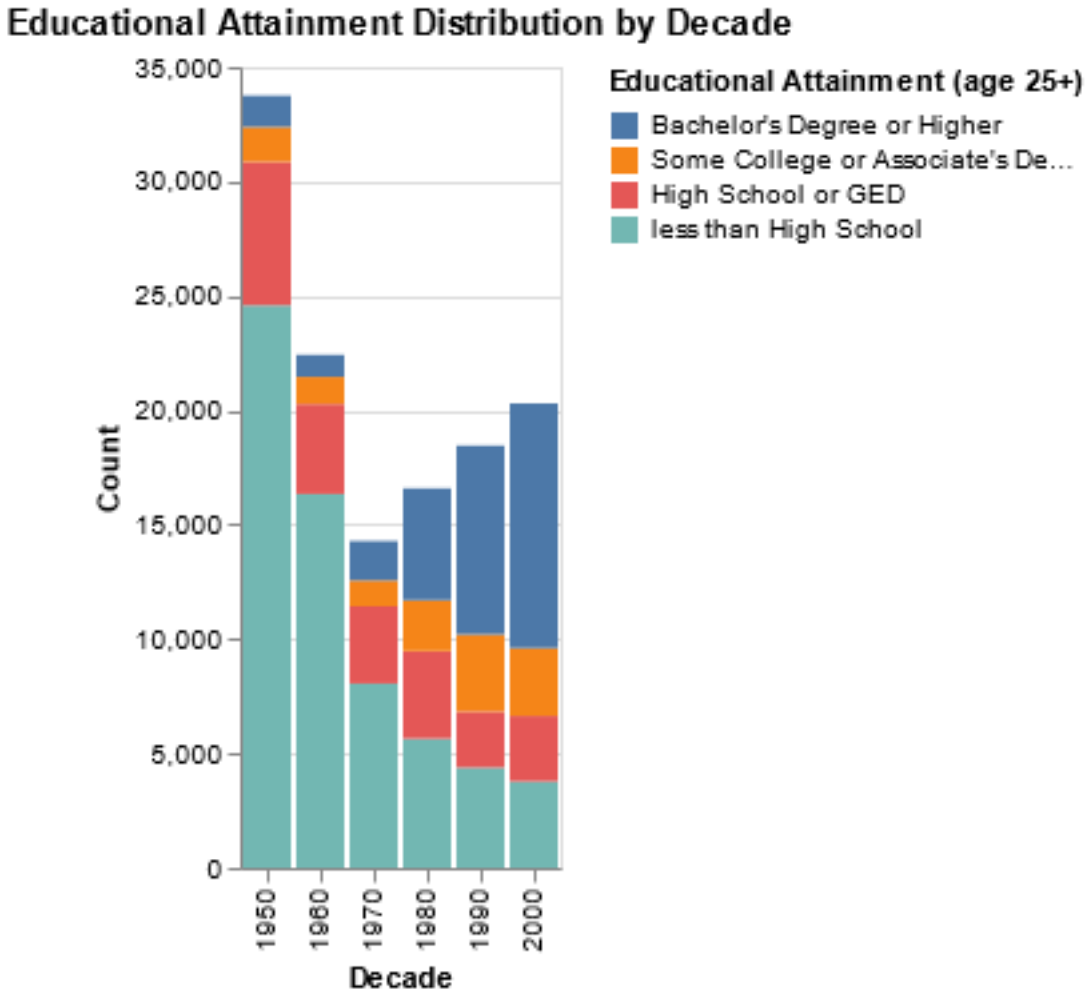



[18]:

```
alt.Chart(df).mark_bar().encode(
    x = 'Decade:O',
    y = 'Count:Q',
    color = alt.Color('Subcategory:N', title = eduField, sort = eduSortOrder),
    order = alt.Order(
        'eduOrdering:N',
        sort = 'ascending'
    )
).transform_filter(
    datum['Category'] == eduField
).transform_calculate(
    eduOrdering = '0'
).properties(
```

```
    title = 'Educational Attainment Distribution by Decade'
)
```

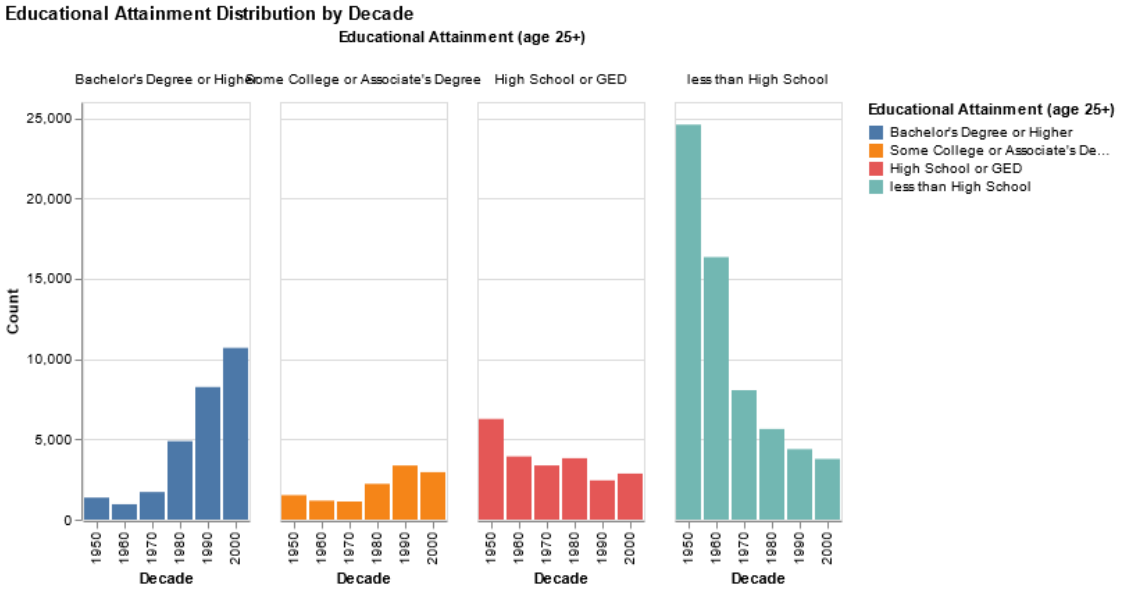
[18]:



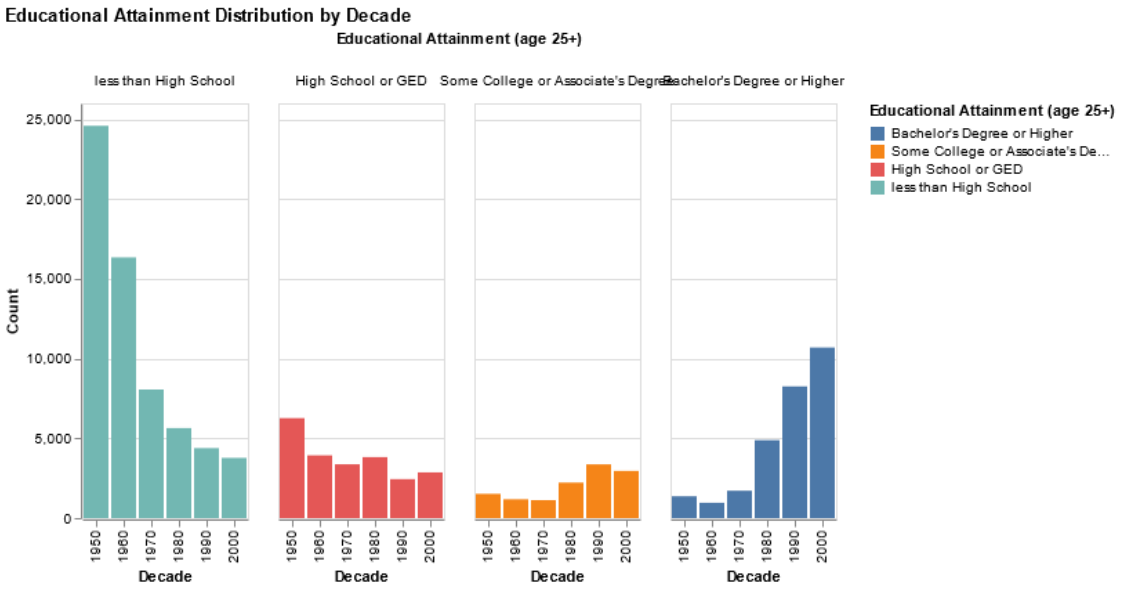

[19]:

```
alt.Chart(df).mark_bar().encode(
    x = 'Decade:O',
    y = 'Count:Q',
    color = alt.Color('Subcategory:N', title = eduField, sort = eduSortOrder),
    column = alt.Column('Subcategory:N', title = eduField, sort = eduSortOrder),
).transform_filter(
    datum['Category'] == eduField
).properties(
    title = 'Educational Attainment Distribution by Decade'
)
```

[19]:



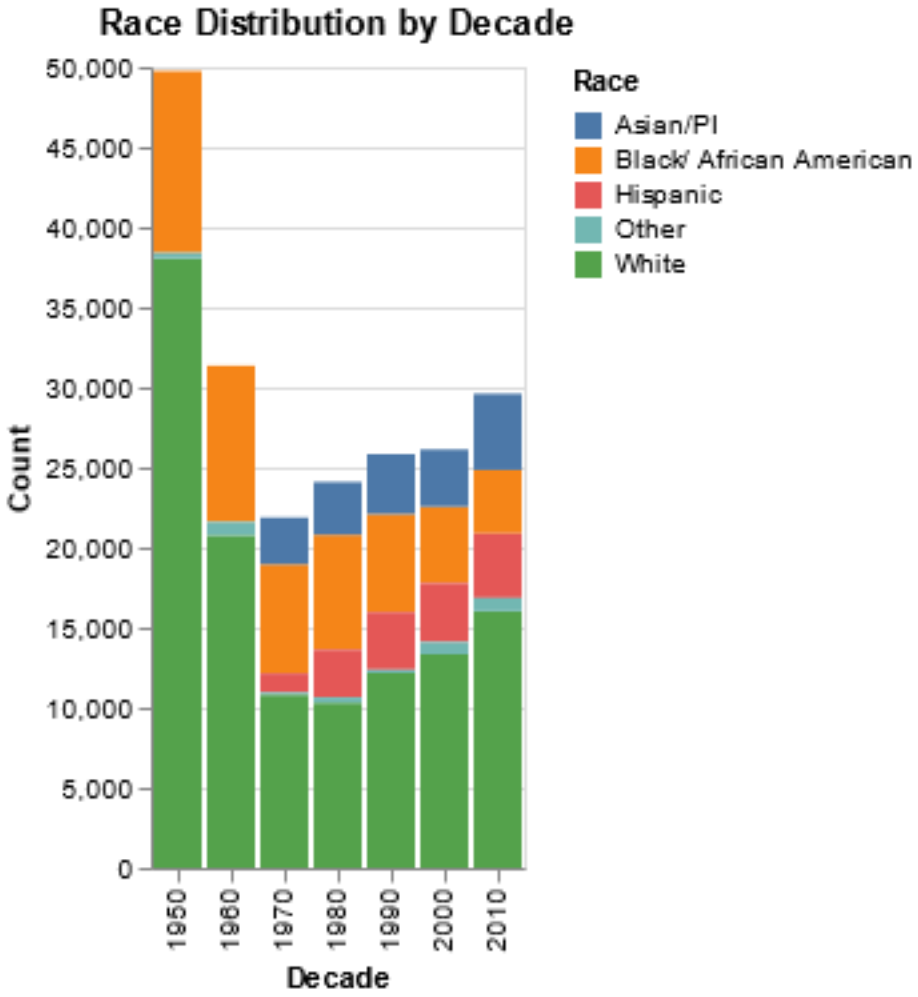

[20]:
```
alt.Chart(df).mark_bar().encode(
    x = 'Decade:O',
    y = 'Count:Q',
    color = alt.Color('Subcategory:N', title = eduField, sort = eduSortOrder),
    column = alt.Column('Subcategory:N', title = eduField, sort = eduSortOrder[:
 ↪:-1]), # ::-1 makes a copy in reverse order
).transform_filter(
    datum['Category'] == eduField
).properties(
    title = 'Educational Attainment Distribution by Decade'
)
```

[20]:


### 5.3 Race by Decade

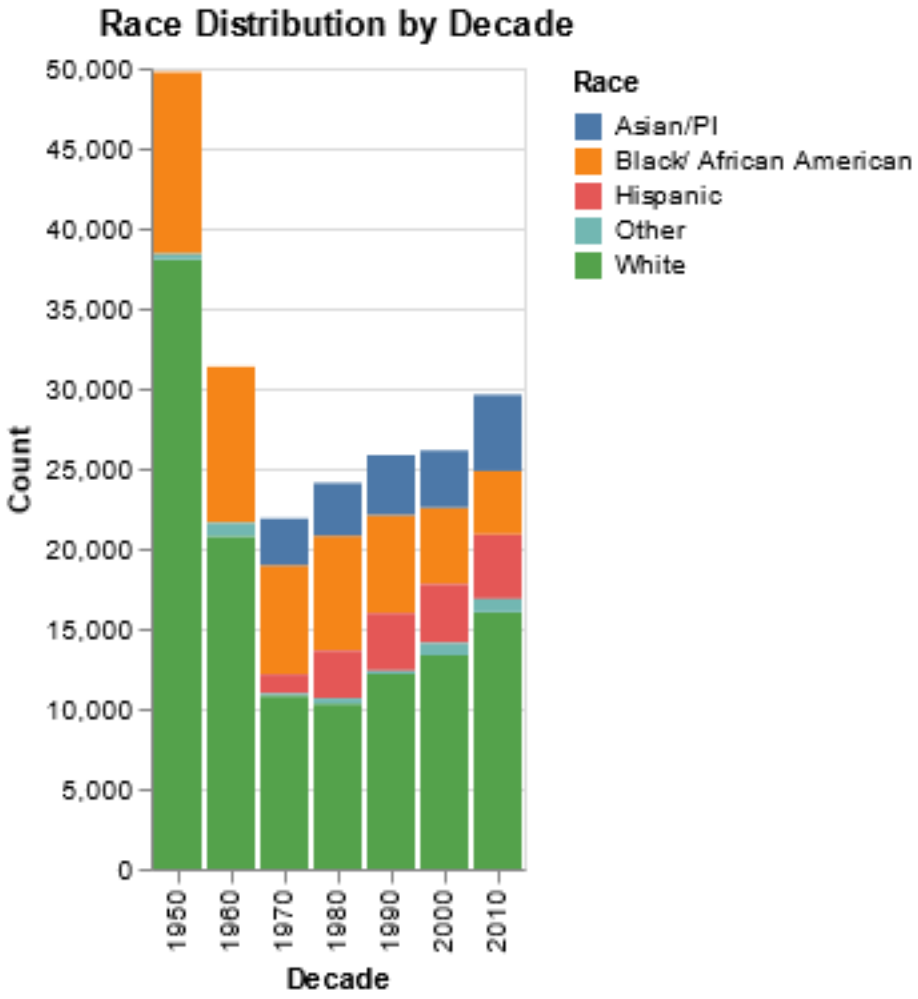
[21]:
```
alt.Chart(df).mark_bar().encode(
    x = 'Decade:O',
    y = 'Count:Q',
    color = alt.Color('Subcategory:N', title = 'Race'),
).transform_filter(
    datum['Category'] == 'Race/ Ethnicity'
).properties(
    title = 'Race Distribution by Decade'
)
```

[21]:



[22]:
```
chart_race = alt.Chart(df).mark_bar().encode(
    x = 'Decade:O',
    y = 'Count:Q',
    color = alt.Color('Subcategory:N', title = 'Race'),
).transform_filter(
    datum['Category'] == 'Race/ Ethnicity'
).properties(
    title = 'Race Distribution by Decade'
)
chart_race
```

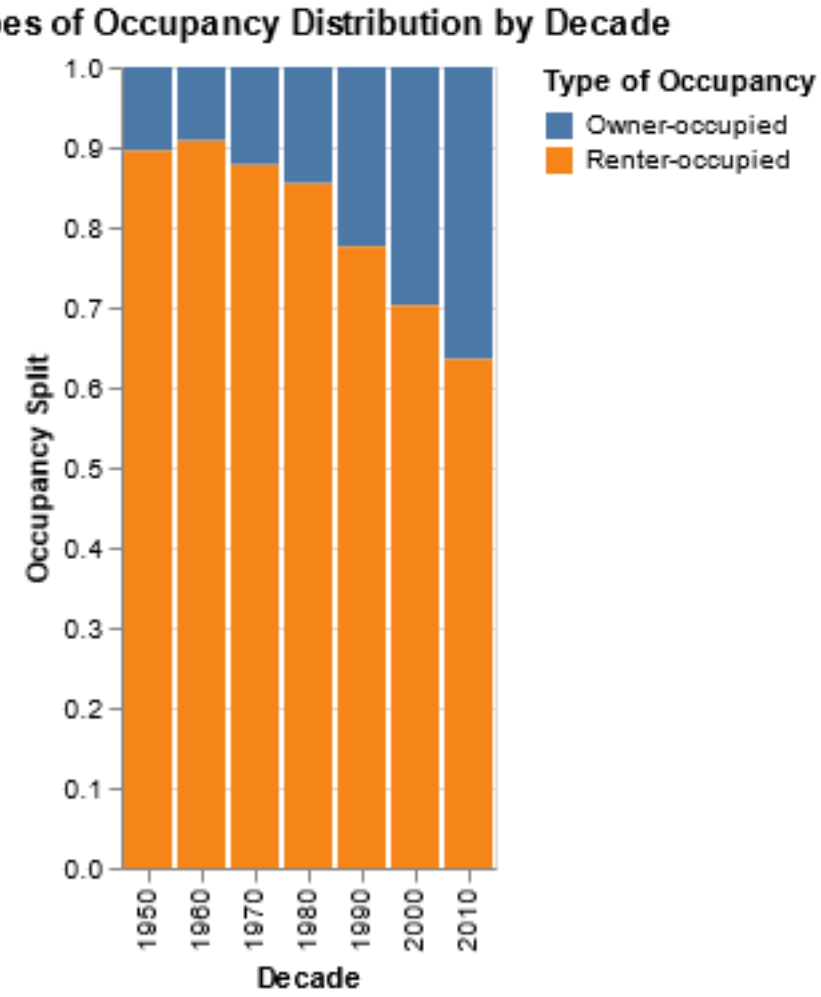
[22]:



### 5.4 Types of Occupancy by Decade

[23]:
```
alt.Chart(df).mark_bar().encode(
    x = 'Decade:O',
    y = alt.Y('Count:Q', stack = 'normalize', title = 'Occupancy Split'),
    color = alt.Color('Subcategory:N', title = 'Type of Occupancy'),
).transform_filter(
    (datum['Category'] == 'Housing Tenure') &
    (datum['Subcategory'] != 'Occupied Housing Units')
).properties(
    title = 'Types of Occupancy Distribution by Decade'
)
```

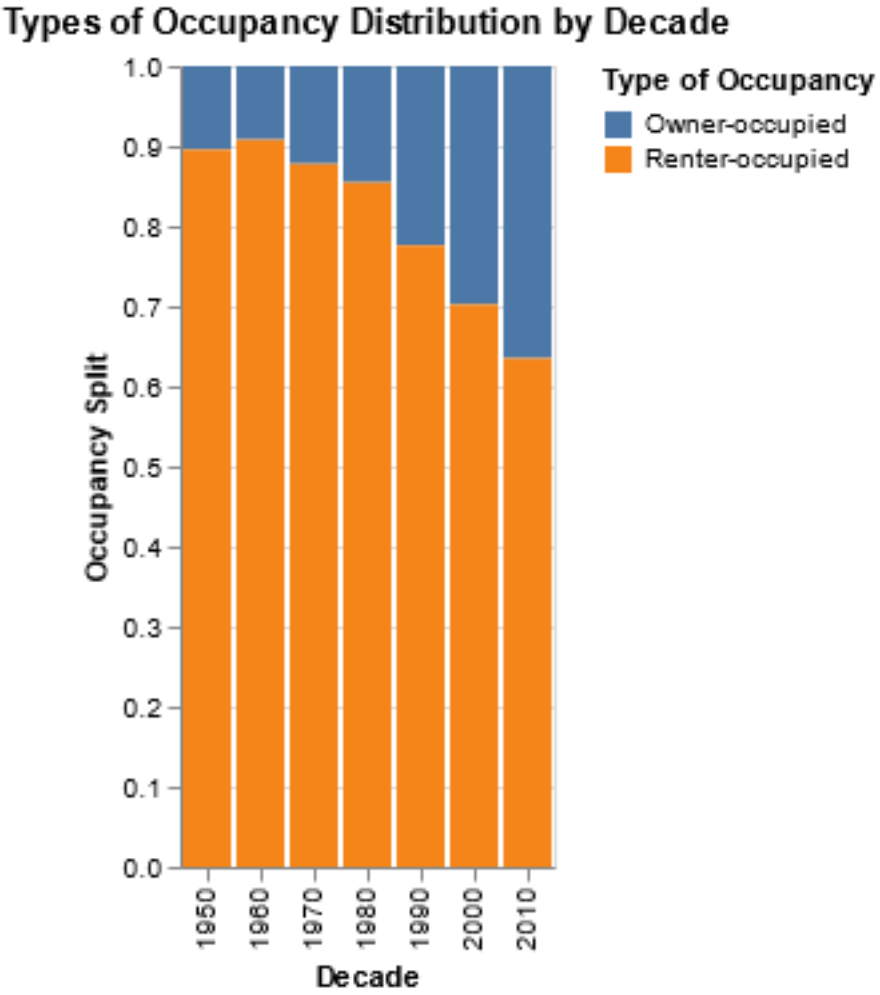
[23]:




[24]:
```
chart_occupancy = alt.Chart(df).mark_bar().encode(
    x = 'Decade:O',
    y = alt.Y('Count:Q', stack = 'normalize', title = 'Occupancy Split'),
```

```
    color = alt.Color('Subcategory:N', title = 'Type of Occupancy'),
).transform_filter(
    (datum['Category'] == 'Housing Tenure') &
    (datum['Subcategory'] != 'Occupied Housing Units')
).properties(
    title = 'Types of Occupancy Distribution by Decade'
)
chart_occupancy
```

[24]:




# 6 Saving Your Results

## 6.1 Saving to a web page...

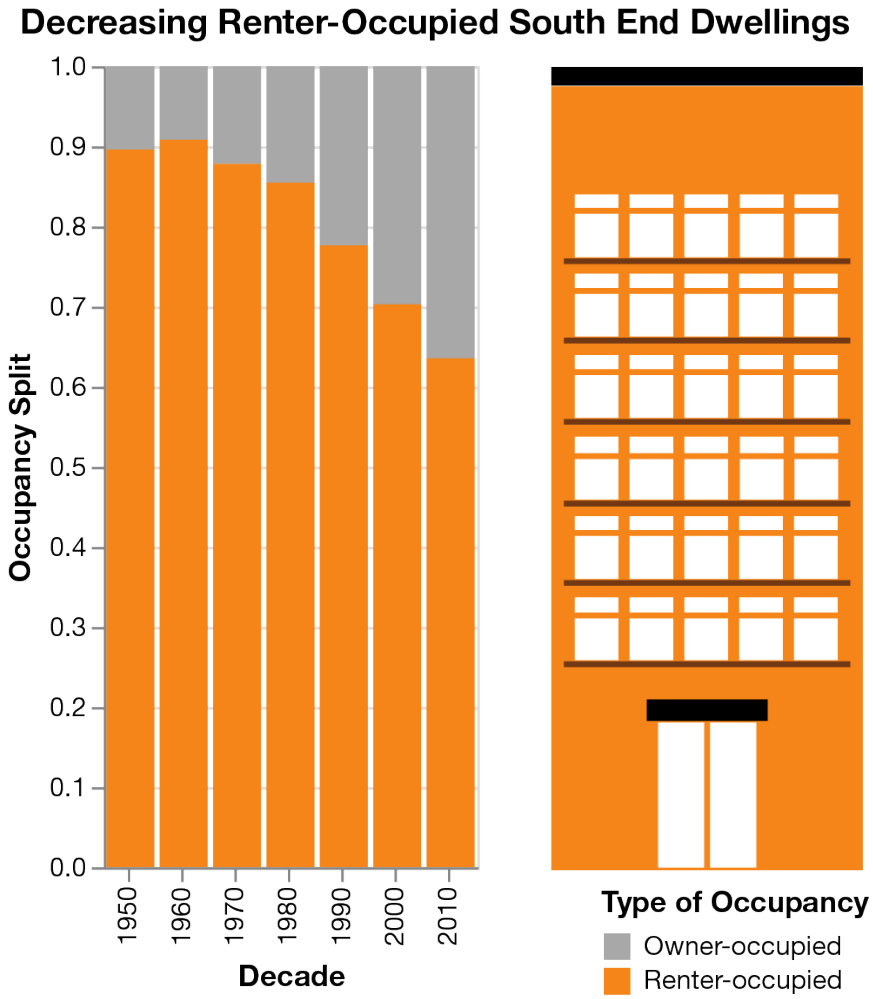
[25]:
```
chart_occupancy.save('chart_occupancy.html', embed_options={'renderer':'svg'})
```

### 6.2 Saving an SVG, PNG, Vega Editor...

**Note:** You do not need to do the following steps yourself. The instructor will simply demonstrate what is possible.

Let's save the chart from above using the ... menu at the top-right in several different formats.

Here is an example infographic you could create by exporting an SVG and editing in Illustrator:




In Illustrator: * Open the SVG. * Use the magic wand tool to select the owner series & the legend. * Edit->Edit Colors->Saturate and reduce saturation. * Add the `apartment_building.svg` and scale to ~20% using Object->Transform->Scale * Select and edit the text at the top to "Decreasing Renter-Occupied South End Dwellings" * File->Export As->SVG