**Final Project- My Business Review of HIV in California**

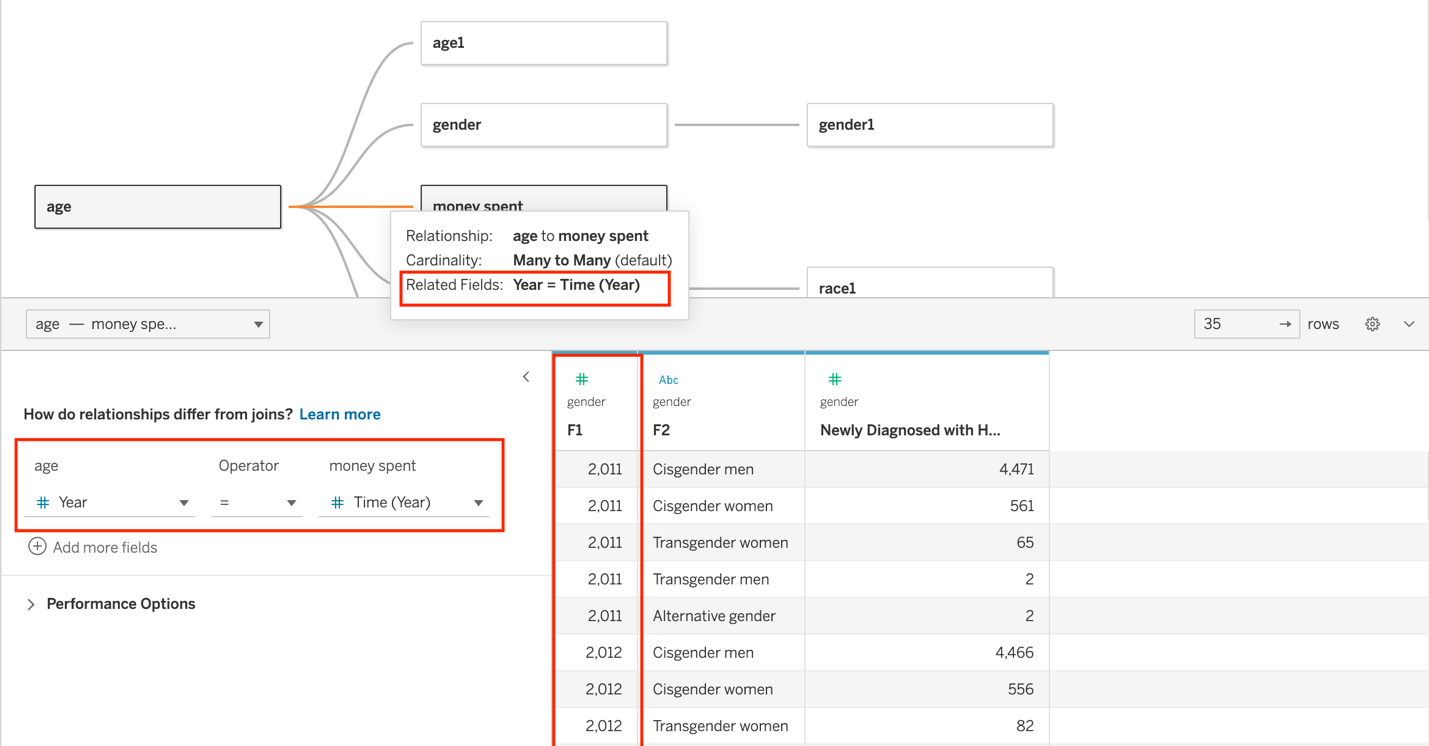
David Tan

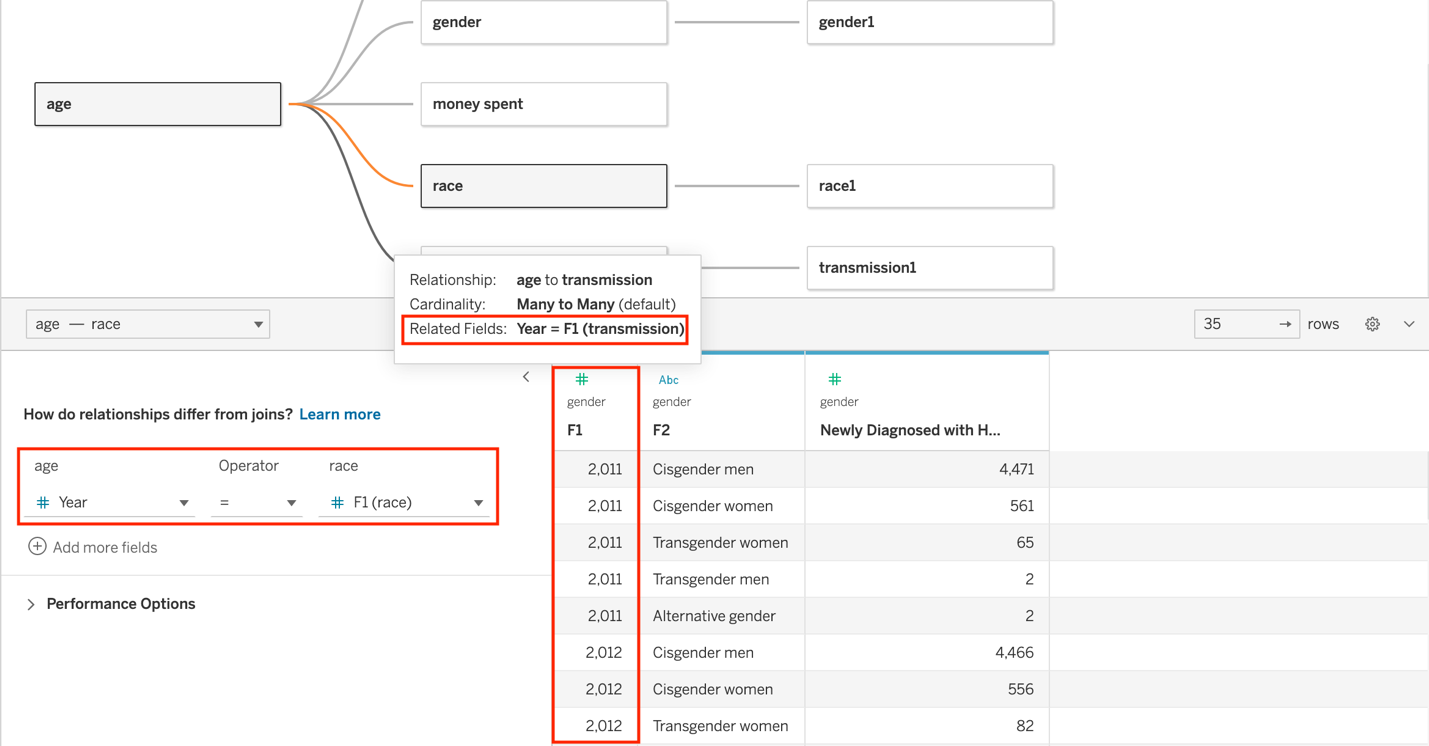
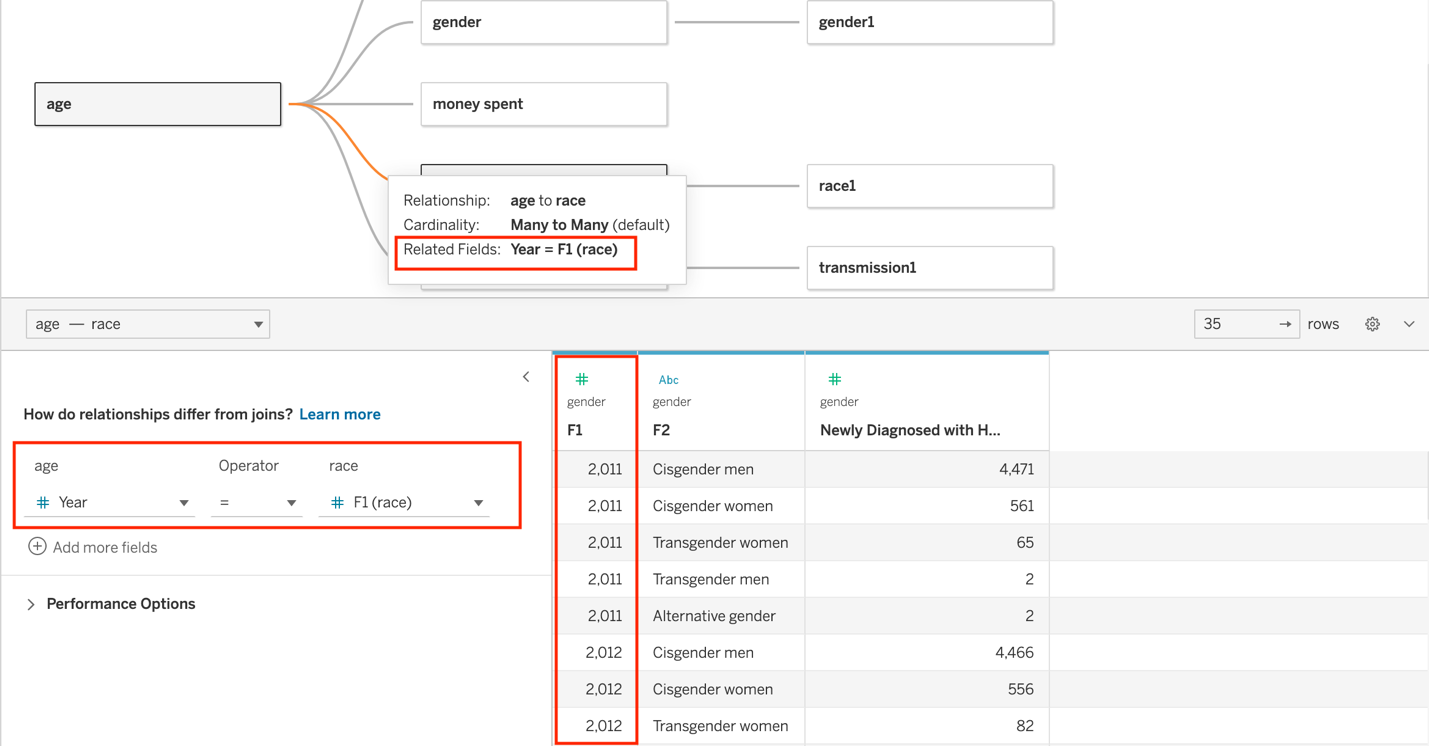
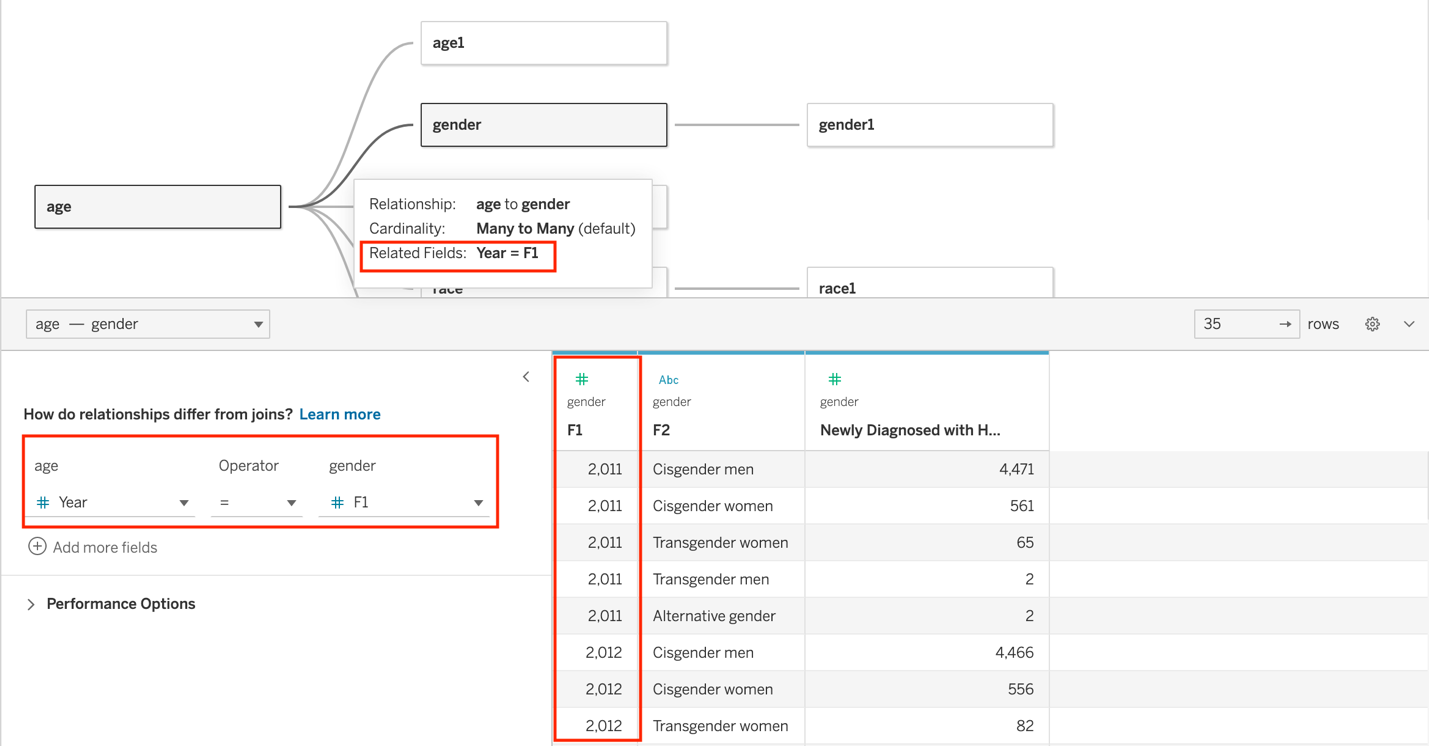
MBA 744- Data Visualization for Business Intelligence

**Table Creating and Joining**

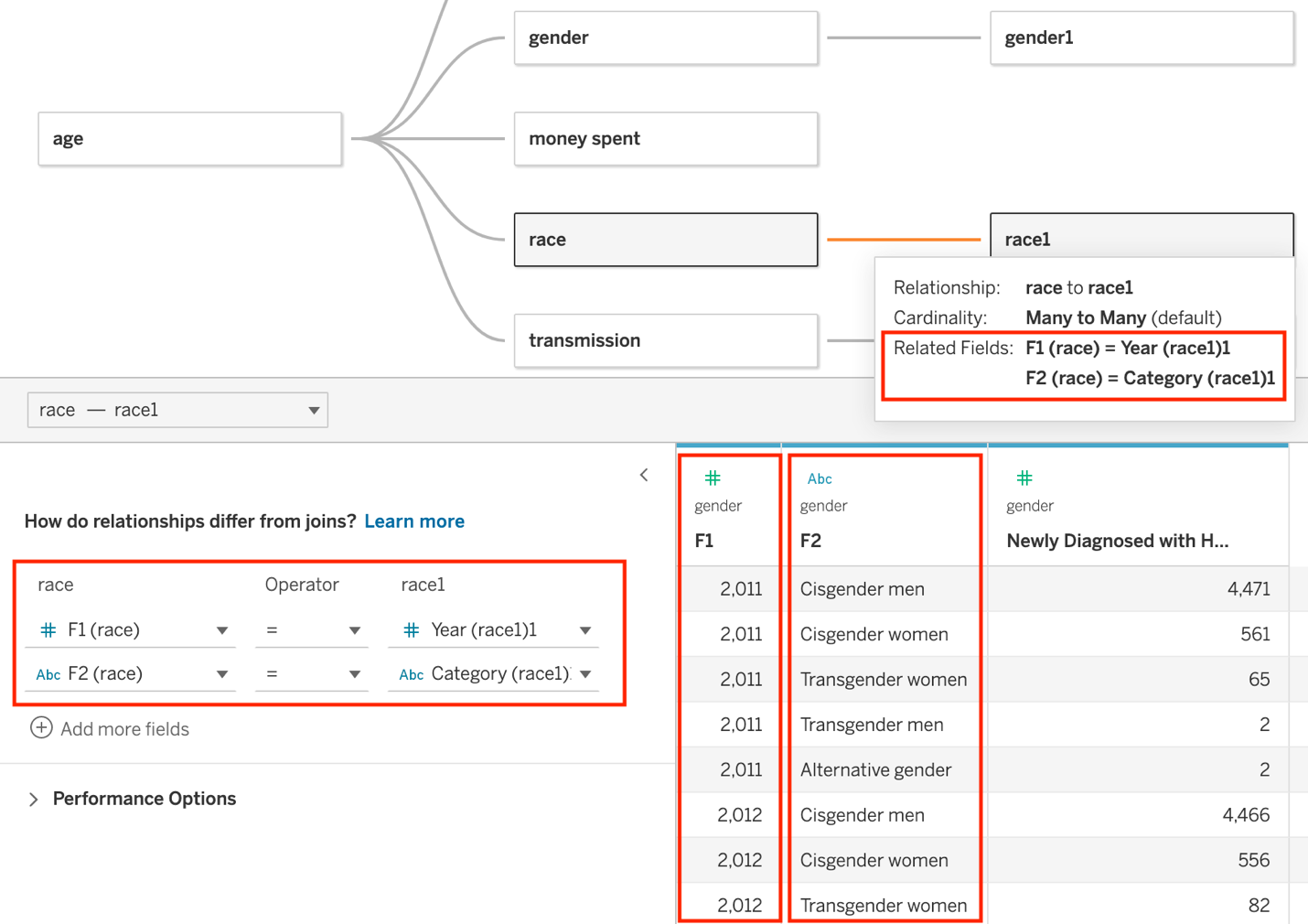
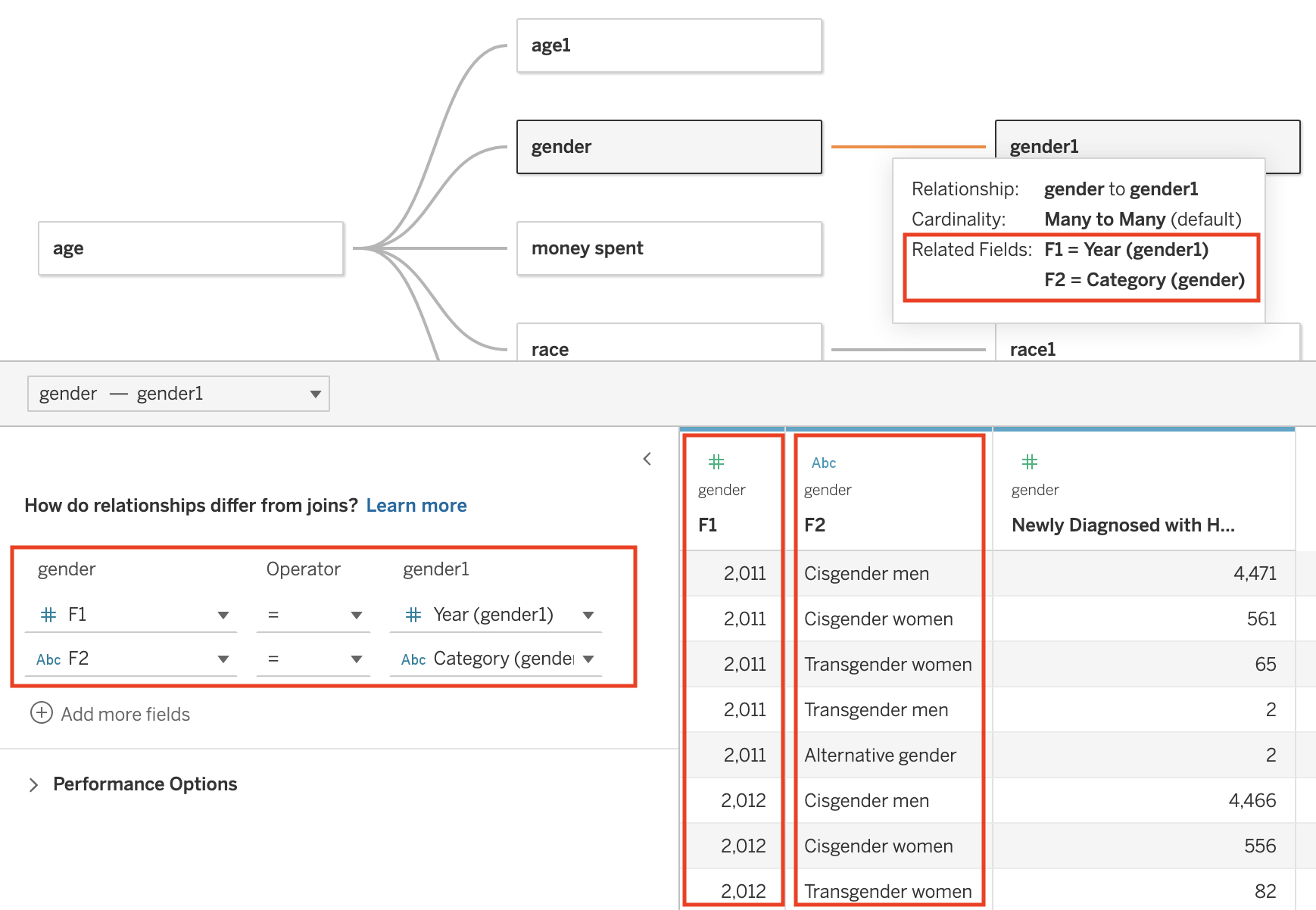
My primary datasets are "excel files" located here: https://data.ca.gov/dataset/hiv-aids-cases. One dataset contains details of persons newly diagnosed with HIV. The second dataset contains information about persons living with HIV/AIDS. These datasets are publicly available and pull additional information from the following data location: California Department of Public Health. The data containing California’s HIV/AIDS funding through the years 2011 to 2021 were obtain here: <https://report.nih.gov/funding/categorical-spending#/>. The data were merged into the first dataset during the process of data cleaning.

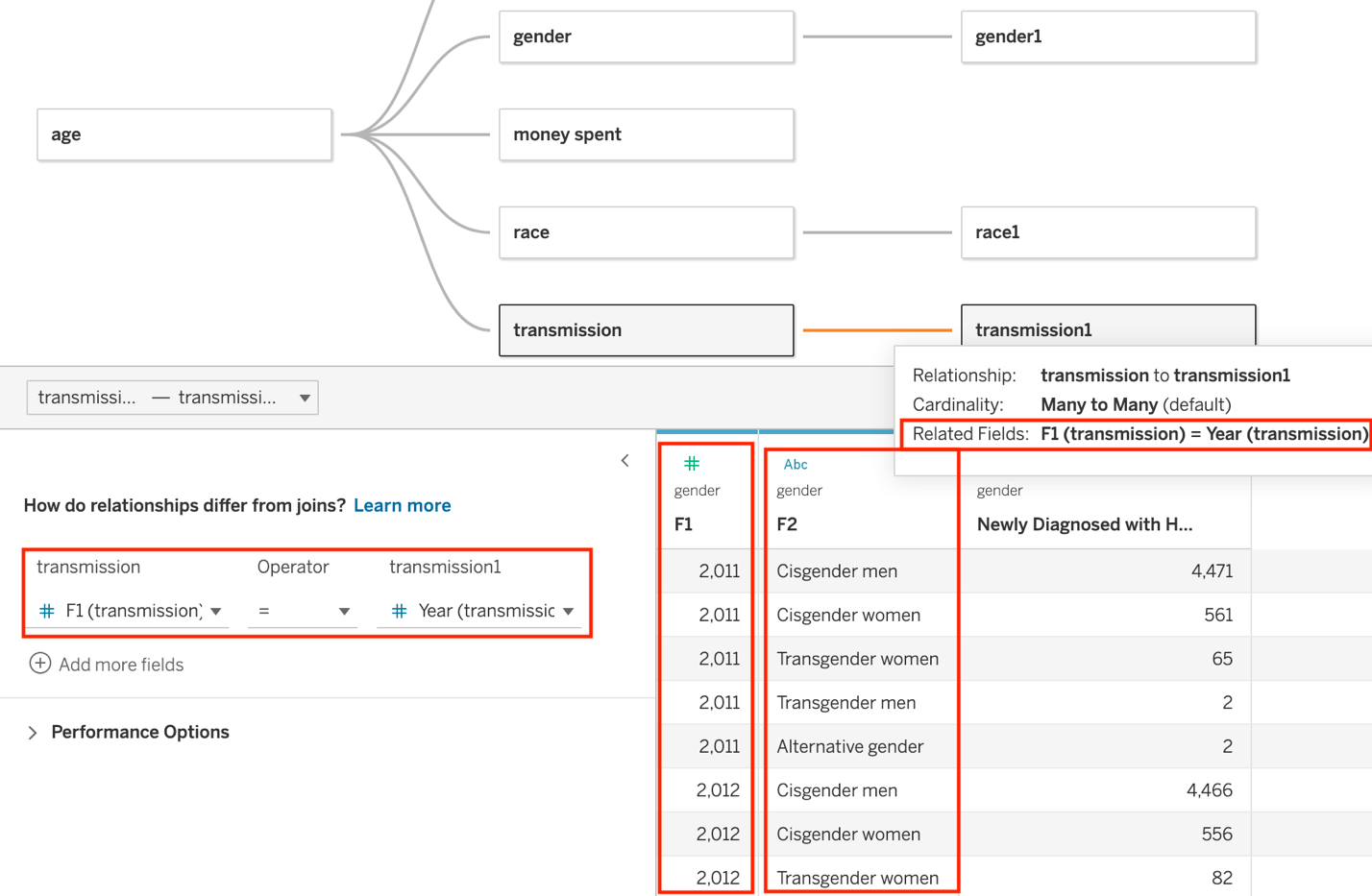
The table Age from the first dataset was used to join the other four tables from the first dataset, i.e., Gender, Money Spent, Race, and Transmission, to itself. This step was necessary as the tables corresponded to separate sheets in the excel files as a result of the data cleaning process. They were joined by using “Year = (their corresponding column for year)” related fields.





Next, the two datasets were joined by using related fields. They were joined by using “Year = (their corresponding column for year)” related fields and by using “F2 = (their corresponding column for F2)” related fields. Except the tables Transmission from both datasets wer4e joined by using only “Year = (their corresponding column for year)” related fields due to the complexity of their F2/category columns.





**Data Cleaning**

The data cleaning process began with the excel files I downloaded from the California Open Data Portal website. The data in the excel files were all jumbled together, and when I initially connected the data to Tableau, the tables and variables were not established, and the data was not useable for the purpose of building visualizations. Therefore, I separated and filtered the data from each dataset into specific sheets, age, gender, race, and transmission.



Furthermore, in order to answer the business question of this project, I merged the data from the website <https://report.nih.gov/funding/categorical-spending#> into the first dataset. Therefore, in the case of the first dataset, there was an extra sheet in the excel file labeled money spent.



Once the data were separated and filtered into their corresponding sheets in the excel files, they were connected to Tableau, where they were organized into their corresponding tables.

Table

Description automatically generated with medium confidence

**Key Performance Indicators**

The business research questions being addressed is how the state of California is spending on the HIV epidemic and if the funds are allocated properly. Furthermore, if HIV is on the rise, I would also like to investigate to see if there is a specific gender, race/ethnicity, age group, and transmission category that has significantly higher HIV/AIDS cases in comparison to the others.

According to the Harvard Business Review, “AIDS is destroying the twin rationales of globalization strategy: cheap labor and fast-growing markets,” and “If your company does business in a developing country anywhere in the world, be it Russia, China, South Africa, or Brazil, AIDS is your business” (Rosen et al., 2014). Therefore, due to the infamous effects that HIV and AIDS have on developing economies and on globalization, it is of great importance to investigate how much the state of California is spending on the epidemic and if the funds are allocated properly. The state of California was chosen due to it containing the largest sub-national economy in the world (Winkler, 2022). The two datasets are joined with a common related field, and in order to answer the business research questions and to provide useful information for businesses in the state of California, the KPI’s being measured are the following:

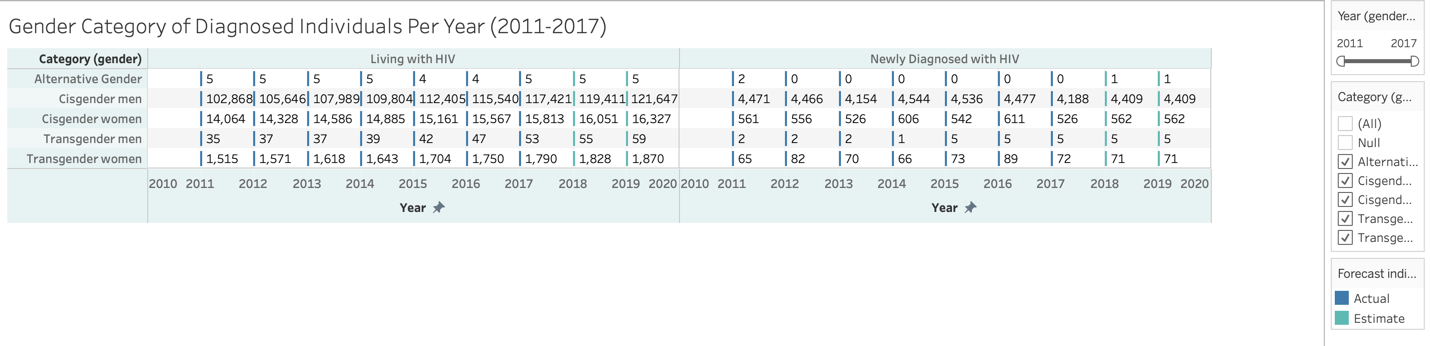
* KPI #1- Age Group: This metric shows the age groups associated with current and new cases of HIV/AIDS. It served the purpose of allowing the viewer to compare the age groups of the individuals newly and currently living with HIV/AIDS in order for key components of this comparison to be identified in order to assist the state of California in properly distributing state funds each consecutive year. Furthermore, this metric was summarized using a side-by-side bar chart, which is depicted below.

Chart, bar chart

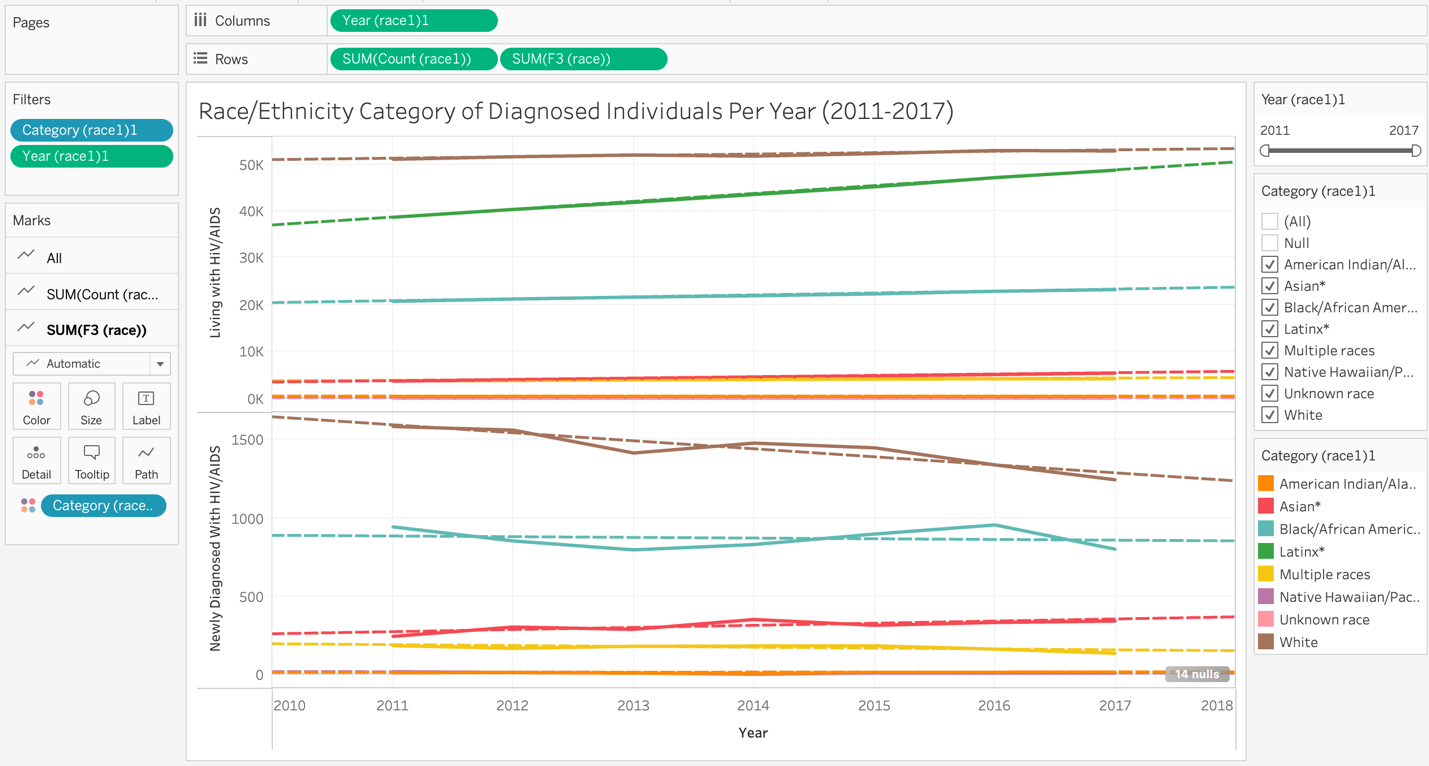
Description automatically generated

A blue color gradient was selected in order to present the data in a clean and concise manner and to highlight and accentuate the information. Three filters were used in order to allow the viewer to dissect the information as desired.

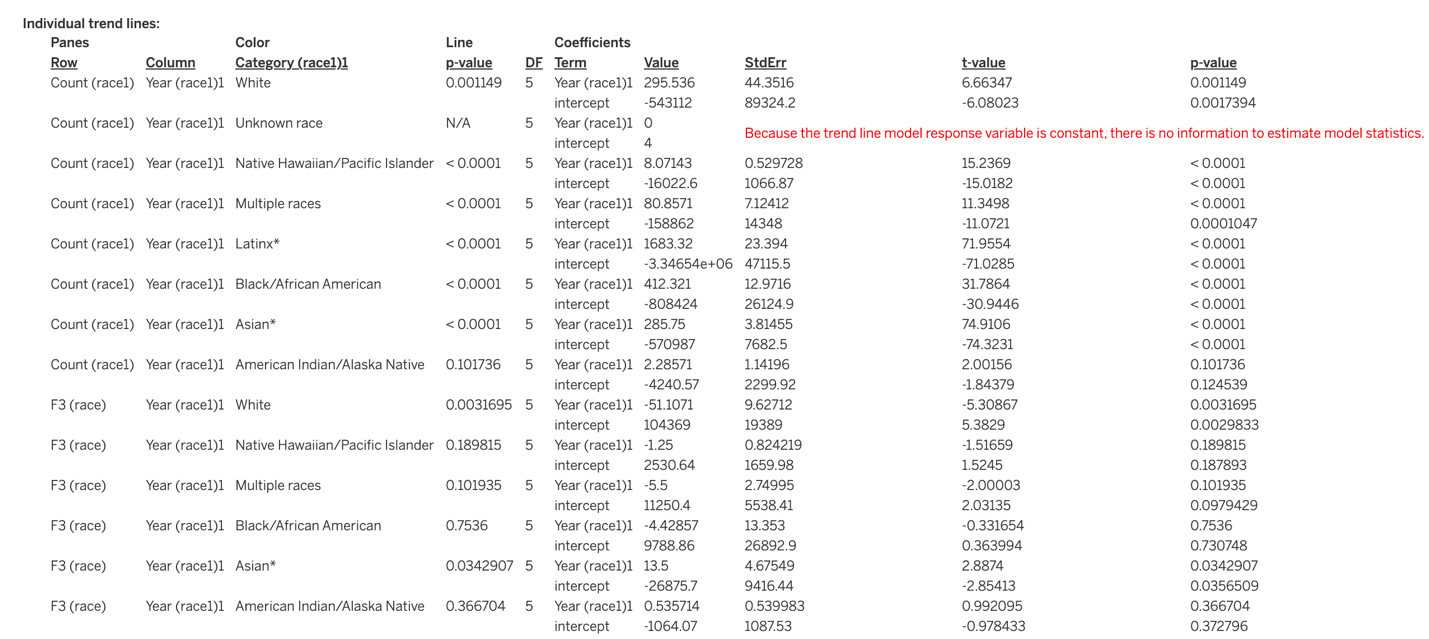
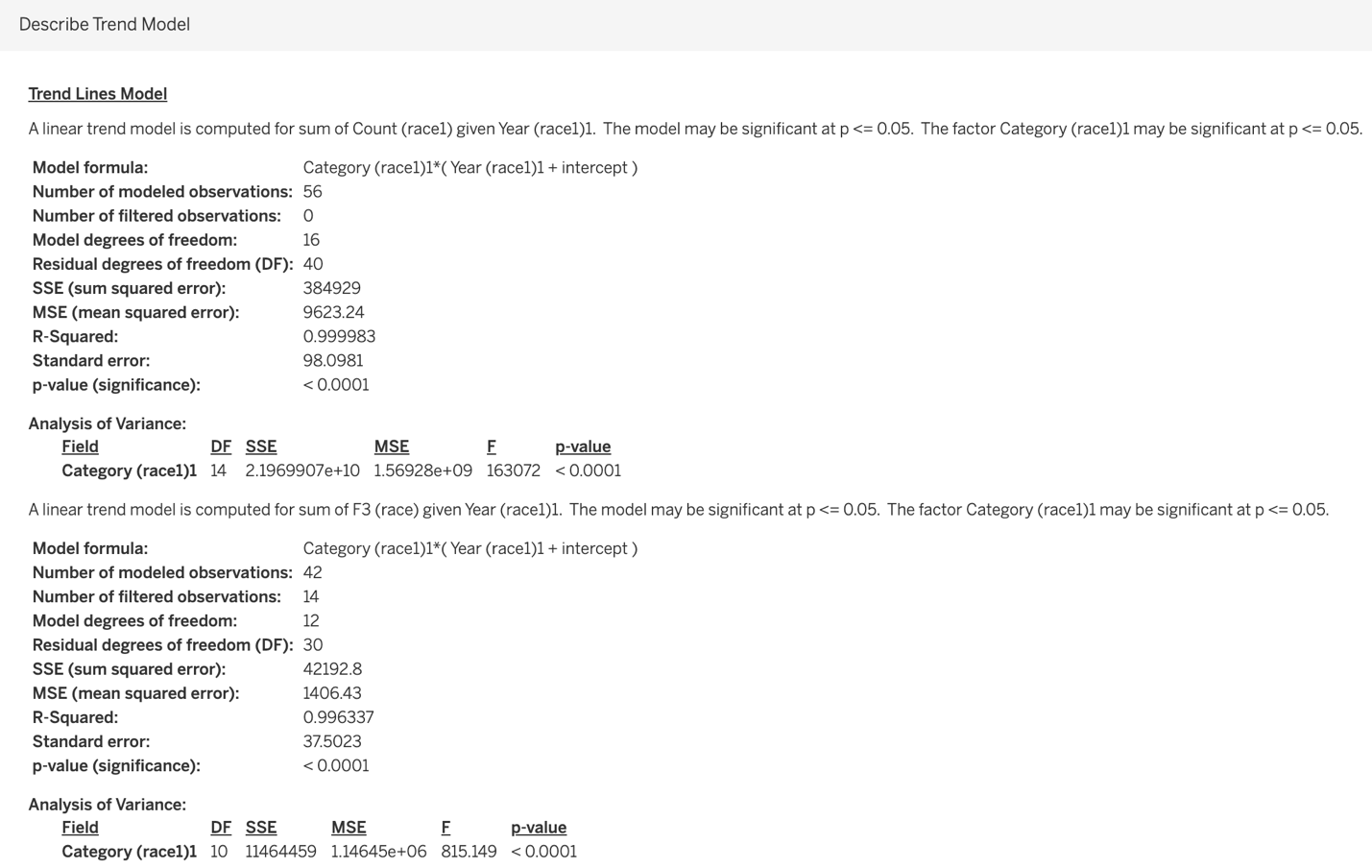
* KPI #2- Gender: This metric shows the genders associated with current and new cases of HIV/AIDS. It served the purpose of allowing the viewer to compare the genders of the individuals newly and currently living with HIV/AIDS in order for key components of this comparison to be identified in order to assist the state of California in properly distributing state funds each consecutive year. Furthermore, this metric was summarized using a text table, which is depicted below. Color was applied to the data table in order to make it eye-catching to the view and to highlight and accentuate the information. The font of the content was made bold in order to highlight and accentuate the information as well. Three filters were used in order to allow the viewer to dissect the information as desired. Forecasting was also applied in order to provide the analyzer with useful information.



* KPI #3- Race/Ethnicity: This metric shows the races/ethnicities associated with current and new cases of HIV/AIDS. It served the purpose of allowing the viewer to compare the race/ethnicity of the individuals newly and currently living with HIV/AIDS in order for key components of this comparison to be identified in order to assist the state of California in properly distributing state funds each consecutive year. Furthermore, this metric was summarized using a side-by-side line graph, which is depicted below.



Color was applied to the data table in order to make it eye-catching to the view and to highlight and accentuate the information. It allows the viewer to analysis the similarities and differences between the two datasets. Two filters were used in order to allow the viewer to dissect the information as desired. Additionally, trend lines were applied for this reason as well. The slopes and intercepts of the trend lines are depicted below.



* KPI #4- Transmission Category: This metric shows the transmission category associated with current and new cases of HIV/AIDS. It served the purpose of allowing the viewer to compare the transmission categories of the individuals newly and currently living with HIV/AIDS in order for key components of this comparison to be identified in order to assist the state of California in properly distributing state funds each consecutive year. Furthermore, this metric was summarized using a treemap, which is depicted below.

Chart, treemap chart

Description automatically generated

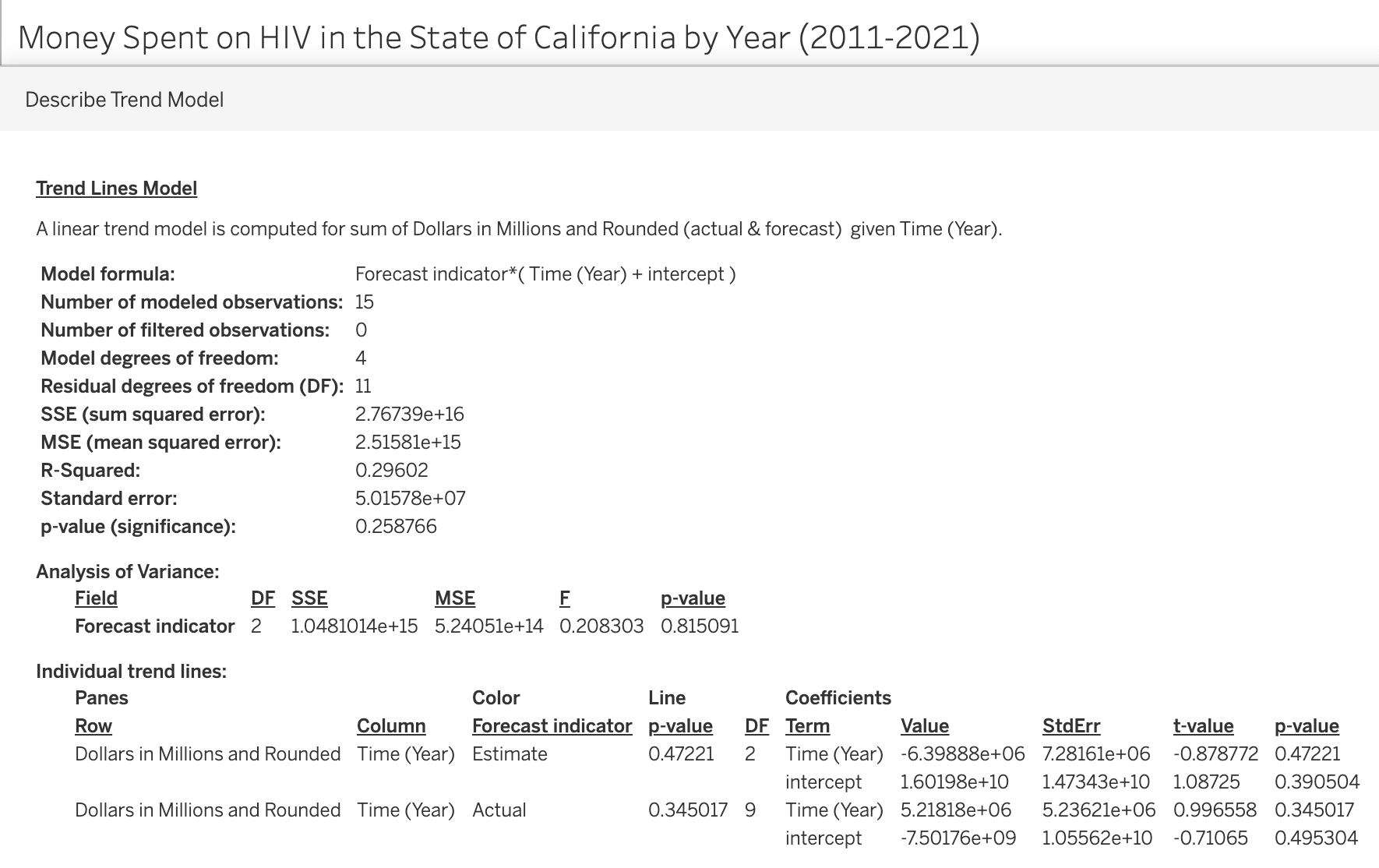
Multiple other visualization methods were attempted on these data, but the treemap model provided the most clean and concise result. Color was applied to the data table in order to make it eye- catching to the view and highlight and accentuate the information. Two filters were used in order to allow the viewer to dissect the information as desired.

* KPI #5- Money spent on HIV in the state of California by year: This metric shows the cases of current and new cases of HIV/AIDS. It served the purpose of allowing the viewer to analyze the money spent on HIV/AIDS in the state of California from 2011-2021 in order for key components of this comparison to be identified in order to assist the state of California in properly distributing state funds each consecutive year. Furthermore, this metric was summarized using a line graph, which is depicted below. Forecasting was also applied in order to provide the analyzer with useful information.

Chart, line chart

Description automatically generated

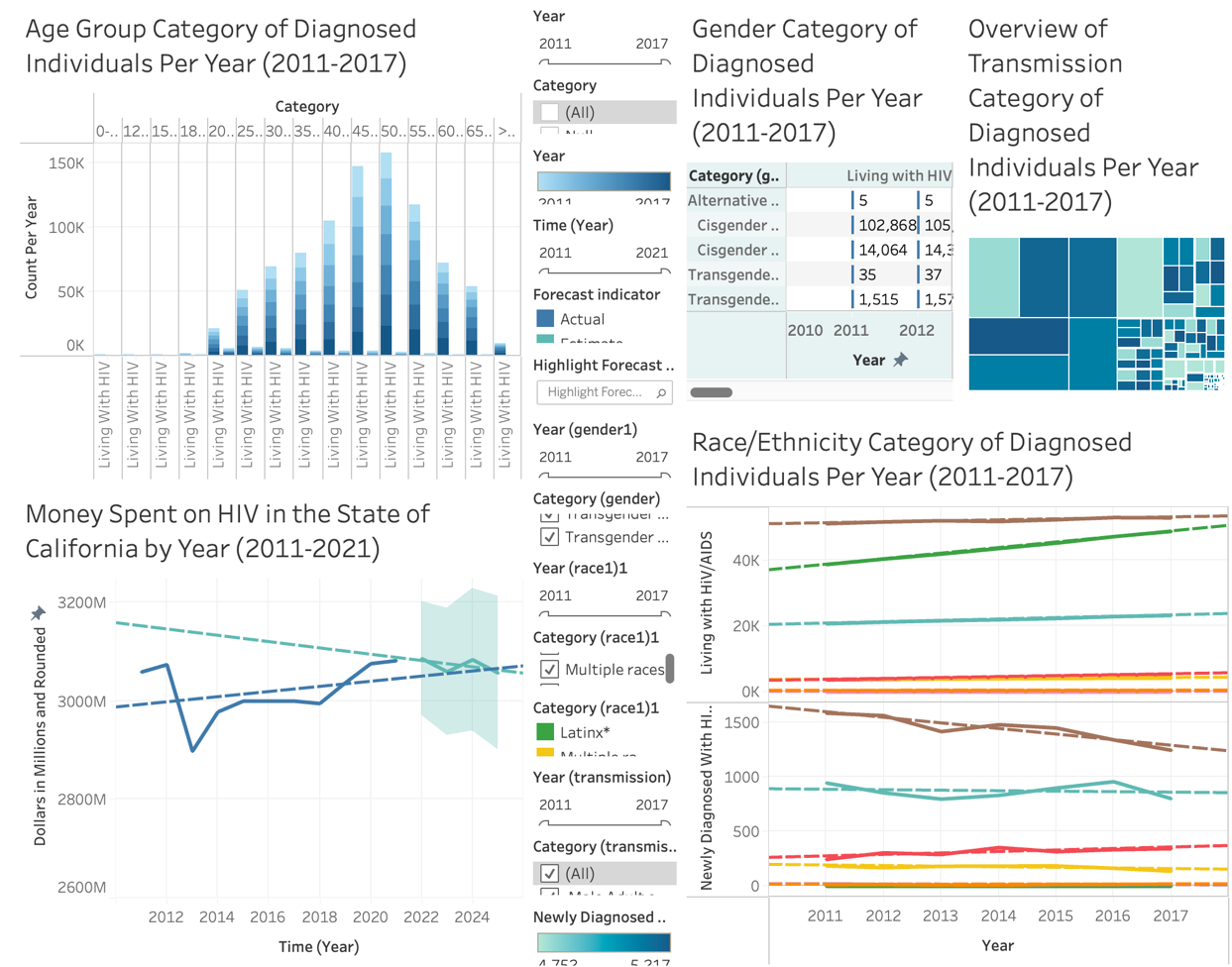
Color was applied to the data table in order to make it eye-catching to the view and highlight and accentuate the information. One filter was used in order to allow the viewer to dissect the information as desired. Additionally, trend lines were applied for this reason as well. The slopes and intercepts of the trend lines are depicted below.



In conclusion, the key data items in the data that I will be using are the age groups, gender, race/ethnicity, transmission categories, and the count of current/ new cases of HIV/AIDs. Furthermore, in order to properly depict these key data variables, the following visualization techniques were used: side-by-side bar chart for KPI #1, a text table for KPI #2, line graphs for KPI #3 and #5, and a treemap for KPI #4. In regard to the targeted audience and the KPI’s being measured, I started thinking more about a dashboard that could accommodate a particular audience for a specific purpose. So, I decided to focus this project on one main audience. The main audience is the state of California health offices as this project measures KPI’s that can used to identify an underuse or overuse of state funds and to identify afflicted groups in California.

**Dashboard Dilemma**

Therefore, for this project, the story that the data of the dashboard told was that of the status of the HIV/AIDS crisis in the state of California from 2011-2017. Due to my audience being the state of California, the factors measured involved the money spent by the state on the HIV/AIDS crisis and the demographic and categorical features of the afflicted residents of California. By providing the state with this analyzed information, I hope to evoke from my audience the response of urgency and lead them to either stabilize or improve the HIV/AIDS crisis in the state of California by properly allocating state funds. Ultimately, no filters or show/hide buttons were applied to this dashboard, but there were attempts to apply them in order to make the dashboard appear less crowded; however, they were not beneficial to the overall appearance of the dashboard. The full "Business Review of HIV in California” dashboard looked like this:



**Storyboard**

My conclusions of the analysis were that in both datasets white, cisgender men who had male-to-male sexual contact presented a higher HIV prevalence than other races, genders, and transmission categories. In the first dataset which contained information about persons newly diagnosed with HIV, the age group greater than or equal to 75 had a higher HIV prevalence than the other age groups, but in the second dataset which contained information about persons living with HIV/AIDS, the age group 50-54 had a higher HIV prevalence than the other age groups. According to my line graph for my KPI #5, the predicted trend of state funds spent on HIV in California is a negative trend line, but the actual trend of state funds spent on HIV in California shows an positive trend line.

For future analysis, I would recommend analyzers to select different datasets in order to avoid the extensive steps I conducted in table creating/joining and data cleaning. I believe by selecting different datasets they will encounter less issues and more accurate results in Tableau as the tables might been readily established. In conclusion, the full "My Business Review of HIV in California” storyboard looked like this:

**Chart

Description automatically generated with medium confidence**

**Link to Tableau Public:**

<https://public.tableau.com/views/finalprojectreport-MyBusinessReviewofHIVinCalifornina/MyBusinessReviewofHIVinCalifornia?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link>

**Link to Video Presentation:**

https://youtu.be/ujVg\_18Br20

References

Rosen, S., Simon, J., Vincent, J. R., MacLeod, W., Fox, M., & Thea, D. M. (2014, August 1). *AIDS is your business*. Harvard Business Review. Retrieved January 29, 2023, from <https://hbr.org/2003/02/aids-is-your-business>.

Winkler, M. A. (2022, October 24). *California poised to overtake Germany as World's no. 4 economy*. Bloomberg.com. Retrieved January 29, 2023, from <https://www.bloomberg.com/opinion/articles/2022-10-24/california-poised-to-overtake-germany-as-world-s-no-4-economy>.