|  |
| --- |
| **Obtaining Statistics for Different Groups**  <NAME OF DATASET> %>%  group\_by(<NAME OF CATEGORICAL VARIABLE>) %>%  summarize(<NAME OF DATASET>,  <NAME OF STAT> = <STAT FUNCTION>(<NAME OF VARIABLE>)  )  For example, to calculate the mean dep\_delay for each airline carrier (from the nycflights dataset) we have:  nycflights %>%  group\_by(carrier) %>%  summarize(mean\_dep\_delay = mean(dep\_delay)  ) |
| **Filtering to Include Specific Values of a Categorical Variable**  filter(nycflights,   <NAME OF CATEGORICAL VARIABLE> %in% c(“<VALUE1>”,“<VALUE2>”,“<VALUE3>”))  The %in% is the “inclusion” operator -- it tells R you are interested in keeping all the values included in the provided list.  The c() function creates a list of levels you are interested in keeping in the dataset. |
| **Side-by-Side Boxplots**  ggplot(data = <NAME OF DATASET>,  mapping = aes(x = <NAME OF VARIABLE>)) +  geom\_boxplot() +  labs(x = "<TITLE FOR THE X-AXIS>")  ***Note:*** This stacks the boxplots horizontally. If you want for your boxplots to be stacked vertically, you use **y =** instead of **x =** . Keep in mind you will need to change the location of you axis label, too! |

|  |
| --- |
| **Faceted Histograms**  ggplot(data = <NAME OF DATASET>,  mapping = aes(x = <NAME OF NUMERICAL VARIABLE>)) +  geom\_histogram(binwidth = <WIDTH OF BINS>) +  facet\_wrap(~ <NAME OF CATEGORICAL VARIABLE>)  labs(x = "<TITLE FOR THE X-AXIS>") |
| **Ridge Plots**  ggplot(data = <NAME OF DATASET>,  mapping = aes(x = <NAME OF NUMERICAL VARIABLE>,  y = <NAME OF CATEGORICAL VARIABLE>)) +  geom\_density\_ridges()+  labs(x = "<TITLE FOR THE X-AXIS>",  y = “<TITLE FOR THE Y-AXIS>”) |
| **Adding Colors to a Plot**  For any plot, you can add a color by mapping it to a variable (categorical or numerical)!  ggplot(data = <NAME OF DATASET>,  mapping = aes(x = <NAME OF VARIABLE>,   y = <NAME OF VARIABLE>,  color = <NAME OF VARIABLE>))  For example, to calculate the mean dep\_delay for each airline carrier (from the nycflights dataset) we have:  nycflights %>%  group\_by(carrier) %>%  summarize(mean\_dep\_delay = mean(dep\_delay)) |
| **Adding Facets to a Plot**  For any plot, you can add a facet by simply adding (with a +) the following code:  facet\_wrap(~ <NAME OF CATEGORICAL VARIABLE>) |
| **Adding Axis Labels to a Plot**  For any plot, you can add axis labels by simply adding (with a +) the following code:  labs(x = “<NAME OF X-AXIS VARIABLE>”,  y = “<NAME OF Y-AXIS VARIABLE>”)  ***Note:*** If you want to rename the legend title associated with the color variable, you add a y = “<NAME OF LEGEND VARIABLE>”)to the above code. |