* Find all flights that:
  + Flew to Chicago
  + Were operated by United or Delta
  + Departed in summer (July, August, and September)
  + Were delayed by at least 30 min, but made up over 15 minutes in flight
  + Departed between midnight and 12 noon (inclusive)
* Another useful dplyr filtering helper is between(). What does it do? Can you use it to simplify the code needed to answer the previous challenges?
* Find all flights that:
  + Flew to Chicago
  + Were operated by United or Delta
  + Departed in summer (July, August, and September)
  + Were delayed by at least 30 min, but made up over 15 minutes in flight
  + Departed between midnight and 12 noon (inclusive)
* Another useful dplyr filtering helper is between(). What does it do? Can you use it to simplify the code needed to answer the previous challenges?
* Find all flights that:
  + Flew to Chicago
  + Were operated by United or Delta
  + Departed in summer (July, August, and September)
  + Were delayed by at least 30 min, but made up over 15 minutes in flight
  + Departed between midnight and 12 noon (inclusive)
* Another useful dplyr filtering helper is between(). What does it do? Can you use it to simplify the code needed to answer the previous challenges?
* Which flights travelled the longest? What was the destination?
* Which flights travelled the shortest? What was destination?
* Which airline has the greatest number of flights in the top 10 list of most-arrival-delayed flights?
* Compare air\_time with arr\_time - dep\_time. What do you expect to see? What do you see? What do you need to do to fix it?
* Compare dep\_time, sched\_dep\_time, and dep\_delay. How would you expect those three numbers to be related?
* Which flights travelled the longest? What was the destination?
* Which flights travelled the shortest? What was destination?
* Which airline has the greatest number of flights in the top 10 list of most-arrival-delayed flights?
* Compare air\_time with arr\_time - dep\_time. What do you expect to see? What do you see? What do you need to do to fix it?
* Compare dep\_time, sched\_dep\_time, and dep\_delay. How would you expect those three numbers to be related?
* Which flights travelled the longest? What was the destination?
* Which flights travelled the shortest? What was destination?
* Which airline has the greatest number of flights in the top 10 list of most-arrival-delayed flights?
* Compare air\_time with arr\_time - dep\_time. What do you expect to see? What do you see? What do you need to do to fix it?
* Compare dep\_time, sched\_dep\_time, and dep\_delay. How would you expect those three numbers to be related?
* Come up with another approach that will give you the same output as not\_cancelled %>% count(dest) and not\_cancelled %>% count(tailnum, wt = distance) (without using count()).
* Look at the number of cancelled flights per day. Is there a pattern? Is the proportion of cancelled flights related to the average delay?
* Which carrier has the worst delays? Challenge: can you disentangle the effects of bad airports vs. bad carriers? Why/why not? (Hint: think about flights %>% group\_by(carrier, dest) %>% summarise(n()))
* Come up with another approach that will give you the same output as not\_cancelled %>% count(dest) and not\_cancelled %>% count(tailnum, wt = distance) (without using count()).
* Look at the number of cancelled flights per day. Is there a pattern? Is the proportion of cancelled flights related to the average delay?
* Which carrier has the worst delays? Challenge: can you disentangle the effects of bad airports vs. bad carriers? Why/why not? (Hint: think about flights %>% group\_by(carrier, dest) %>% summarise(n()))
* Come up with another approach that will give you the same output as not\_cancelled %>% count(dest) and not\_cancelled %>% count(tailnum, wt = distance) (without using count()).
* Look at the number of cancelled flights per day. Is there a pattern? Is the proportion of cancelled flights related to the average delay?
* Which carrier has the worst delays? Challenge: can you disentangle the effects of bad airports vs. bad carriers? Why/why not? (Hint: think about flights %>% group\_by(carrier, dest) %>% summarise(n()))
* Make a visualization of the top five destinations of each airline out of PDX
* What time of day should you fly if you want to avoid delays as much as possible?
* Look at each destination. Can you find flights that are suspiciously fast? (i.e. flights that represent a potential data entry error). Compute the air time a flight relative to the shortest flight to that destination. Which flights were most delayed in the air?
* Find all destinations that are flown by at least two carriers. Use that information to rank the carriers.
* What is the most commonly flown route in the PNW?
* Which plane (tailnum) has the worst on-time record?
* Make a visualization of the top five destinations of each airline out of PDX
* What time of day should you fly if you want to avoid delays as much as possible?
* Look at each destination. Can you find flights that are suspiciously fast? (i.e. flights that represent a potential data entry error). Compute the air time a flight relative to the shortest flight to that destination. Which flights were most delayed in the air?
* Find all destinations that are flown by at least two carriers. Use that information to rank the carriers.
* What is the most commonly flown route in the PNW?
* Which plane (tailnum) has the worst on-time record?