1. Which destination in the flights database is the furthest distance away?

```
select dest from flights order by distance desc limit 1

--If we need to both destination code and name
select f.dest, b.name
        from flights f inner join airports b on f.dest = b.faa
        order by f.distance desc limit 1
```

dest character(3)	name character varying
HNL	Honolulu Intl

2. What are the different numbers of engines in the planes table? For each number of engines, which aircraft have the most number of seats?

```
select distinct engines from planes order by engines
```

Data Output

	engines integer
1	1
2	2
3	3
4	4

2.2). For each number of engines, which aircraft have the most number of seats?

```
select distinct p.engines, p.manufacturer, p.seats
from ( select engines,max(seats) as maxseats
        from planes group by engines ) as temp
 inner join planes as p on p.engines = temp.engines and p.seats = temp.maxseats
 order by p.engines,p.manufacturer
```

Data Output | Explain | Messages | History

	engines integer	manufacturer character varying	seats integer
1	1	DEHAVILLAND	16
2	2	BOEING	400
3	3	AIRBUS	379
4	4	BOEING	450

3. What weather conditions are associated with New York City departure delays?

```
select f.origin,f.dest,f.flight,f.dep_delay,(f.year || '-' || f.month || '-' || f.day) as
dep_date,f.hour,w.temp,w.dewp,w.humid,w.wind_dir,w.wind_speed, w.wind_gust,
w.precip,w.pressure,w.visib  from flights f join weather w on f.year = w.year and f.month = w.month and
f.day = w.day and f.hour = w.hour and f.origin = w.origin where f.dep_delay > 0 and f.origin in ('JFK',
'LGA', 'EWR') order by dep_date desc, f.origin
```

Output pane

Data Output | Explain | Messages | History

	origin character(3)	dest character(3)	flight integer	dep_delay integer	dep_date text	hour integer	temp double precision	dewp double precision	humid double precision	wind_dir integer	wind_speed double precision	wind_gust double precision	precip double precision	pressure double precision	visib double precision
1	EWR	TPA	1115	4	2013-9-9	6	57.02	42.98	59.31	20	9.20624	10.5943568672	0	1019.9	10
2	EWR	GRR	4570	2	2013-9-9	20	69.98	51.98	52.8	140	9.20624	10.5943568672	0	1022.2	10
3	EWR	LAS	1289	1	2013-9-9	7	57.02	44.06	61.81	10	8.05546	9.2700622588	0	1020.8	10
4	EWR	STL	4672	5	2013-9-9	20	69.98	51.98	52.8	140	9.20624	10.5943568672	0	1022.2	10
5	EWR	DEN	1682	9	2013-9-9	7	57.02	44.06	61.81	10	8.05546	9.2700622588	0	1020.8	10
6	EWR	SFO	754	7	2013-9-9	20	69.98	51.98	52.8	140	9.20624	10.5943568672	0	1022.2	10
7	EWR	SEA	15	1	2013-9-9	7	57.02	44.06	61.81	10	8.05546	9.2700622588	0	1020.8	10
8	EWR	MDW	565	4	2013-9-9	20	69.98	51.98	52.8	140	9.20624	10.5943568672	0	1022.2	10
9	EWR	LAX	1415	14	2013-9-9	10	55.04	44.06	66.41	10	8.05546	9.2700622588	0	1023.1	10
10	EWR	MCI	4321	1	2013-9-9	20	69.98	51.98	52.8	140	9.20624	10.5943568672	0	1022.2	10
11	EWR	DFW	1218	2	2013-9-9	8	55.04	44.06	66.41	20	9.20624	10.5943568672	0	1021.4	10

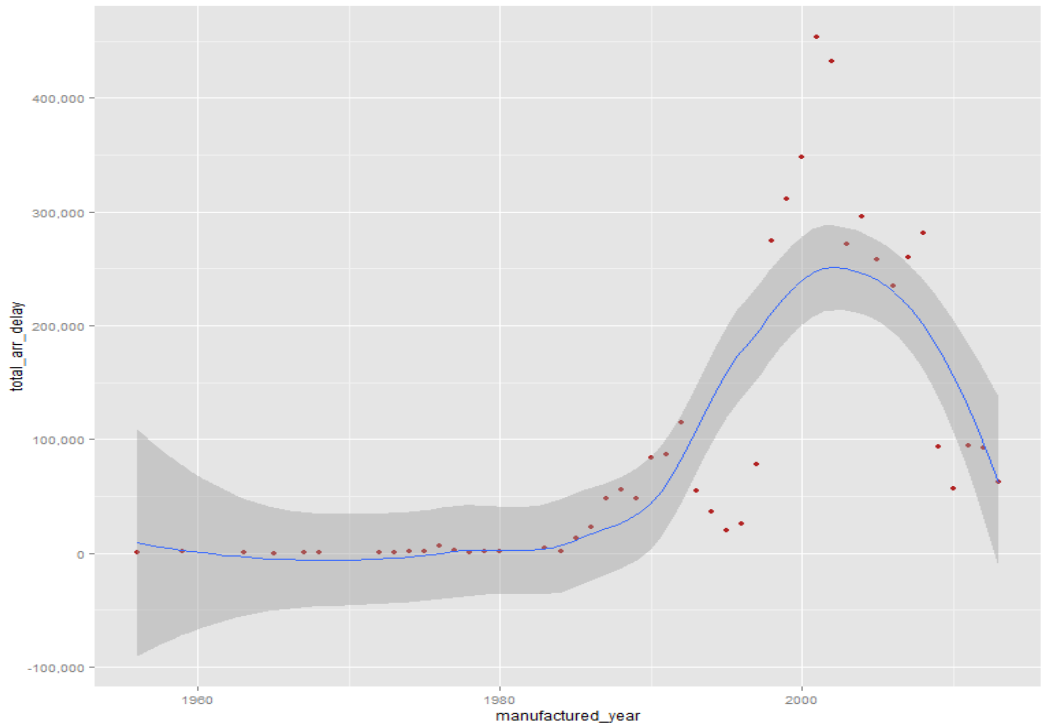
4. Are older planes more likely to be delayed?

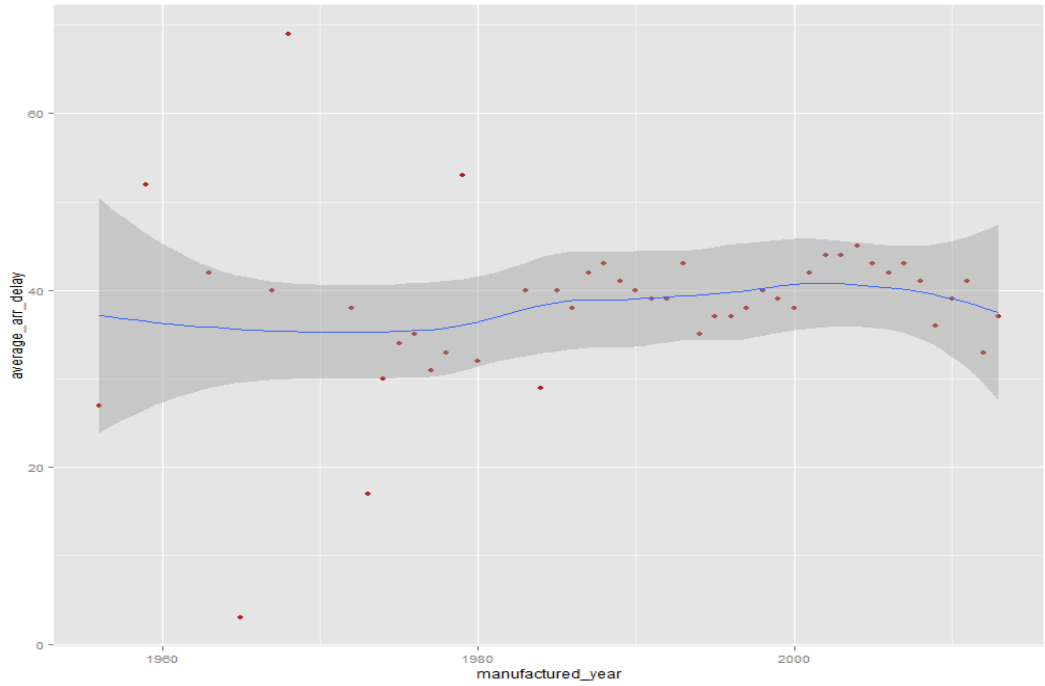
Let's first try with the total delay for each manufactured year. And then with average delay per manufacture year.

```
select p.year as manufacture_year,
        round(sum(f.arr_delay)) as total_arr_delay
from    planes p join flights f on p.tailnum = f.tailnum
where f.arr_delay > 0 and p.year is not null
group by manufacture_year order by manufacture_year desc, total_arr_delay
```

```
select p.year as manufacture_year, round(avg(f.arr_delay)) as average_arr_delay
from    planes p join flights f on p.tailnum = f.tailnum
where f.arr_delay > 0 and p.year is not null
group by manufacture_year order by manufacture_year desc, average_arr_delay
```

```
install.packages('RPostgreSQL')
install.packages('ggplot2')
library(RPostgreSQL)
library(ggplot2)
library(scales)


con <- dbConnect(PostgreSQL(), user= "postgres", password="test123", dbname="flights")
rs_total_delays <- dbSendQuery(con,
                  "select p.year as manufactured_year, round(sum(f.arr_delay)) as total_arr_delay from planes p join flights f on p.tailnum = f.tailnum where
                  f.arr_delay > 0 and p.year is not null  group by manufactured_year order by manufactured_year desc, total_arr_delay")

#fetch observations
df_total_delays <- fetch(rs_total_delays, n=-1)

# Clearing the result set
dbClearResult(rs_total_delays)

#fetch observations
rs_avg_delays <- dbSendQuery(con,
                             "select p.year as manufactured_year, round(avg(f.arr_delay)) as average_arr_delay from planes p join flights f on
                             p.tailnum = f.tailnum where f.arr_delay > 0 and p.year is not null  group by manufactured_year order by manufactured_year desc,
                             average_arr_delay")

df_avg_delays <- fetch(rs_avg_delays, n=-1)

# Clearing the result set
dbClearResult(rs_avg_delays)

# Disconnect
dbDisconnect(con)

# Plot the graph for total delays vs manf year
ggplot(data = df_total_delays, aes(x=manufactured_year, y=total_arr_delay)) + geom_point(color="firebrick") + scale_y_continuous(labels=comma) + stat_smooth()

# Plot the graph for avg delays vs manf year
ggplot(data = df_avg_delays, aes(x=manufactured_year, y=average_arr_delay)) + geom_point(color="firebrick") + scale_y_continuous(labels=comma) + stat_smooth()
```





From the above observations, we can conclude that delays are not caused by older planes

5. Ask (and if possible answer) a question that also requires joining information from two or more tables in the flights database, and/or assumes that additional information can be collected in advance of answering your question

Question -
Identify the top-3 worst and best airlines carriers with respect to arrival delays overall.

Best Airlines (top-3)

```
Select a.carrier,  a.name, round(avg(f.arr_delay)) as avg_arr_delay
from airlines a join flights f   on a.carrier = f.carrier
 and f.arr_delay >= 0            group by a.carrier,  a.name order by avg_arr_delay asc limit 3
```

carrier character(2)	name character varying	avg_arr_delay numeric
US	US Airways Inc.	29
AS	Alaska Airlines	34
HA	Hawaiian Airlir	35

Worst Airlines (top-3)

```
select a.carrier,a.name,round(avg(f.arr_delay)) as avg_arr_delay from airlines a
join flights f on a.carrier = f.carrier and f.arr_delay >= 0 group by a.carrier, a.name order by
avg_arr_delay desc limit 3
```

carrier character(2)	name character varying	avg_arr_delay numeric
OO	SkyWest Airline	61
YV	Mesa Airlines I	50
9E	Endeavor Air In	48