

GDA of *AirCrash* (from **vcdExtra**)

Background Data on all fatal commercial airplane crashes from 1993–2015. Excludes small planes (less than 6 passengers) and non-commercial (cargo, military, private) aircraft.

Aims What are the trends in fatal airplane crashes? At what stage of a flight do crashes occur? What are the reasons for the crashes and have they varied by stage of a flight and over time?

Source David McCandless, who provides a more detailed dataset (<http://www.informationisbeautiful.net/visualizations/plane-truth-every-single-commercial-plane-crash-visualized/>)

Structure 439 observations on 5 variables (1 numeric, 2 factors, date, and year)

First consider the frequency of crashes over the 23 years of the dataset. Bear in mind that numbers of flights have increased over the period and that the data for 2015 are only up till the end of March of that year.

```
library(dplyr); library(gridExtra)
data(AirCrash, package="vcdExtra")
AirCrash <- AirCrash %>% mutate(year=factor(Year))
a1 <- ggplot(AirCrash, aes(reorder(year, -Year))) + geom_bar() +
  coord_flip() + xlab("") + ylab("Crashes per year")
a2 <- ggplot(AirCrash, aes(reorder(year, -Year))) + geom_bar(aes(weight=Fatalities)) +
  coord_flip() + xlab("") + ylab("Fatalities per year")
grid.arrange(a1, a2, ncol=2)
```

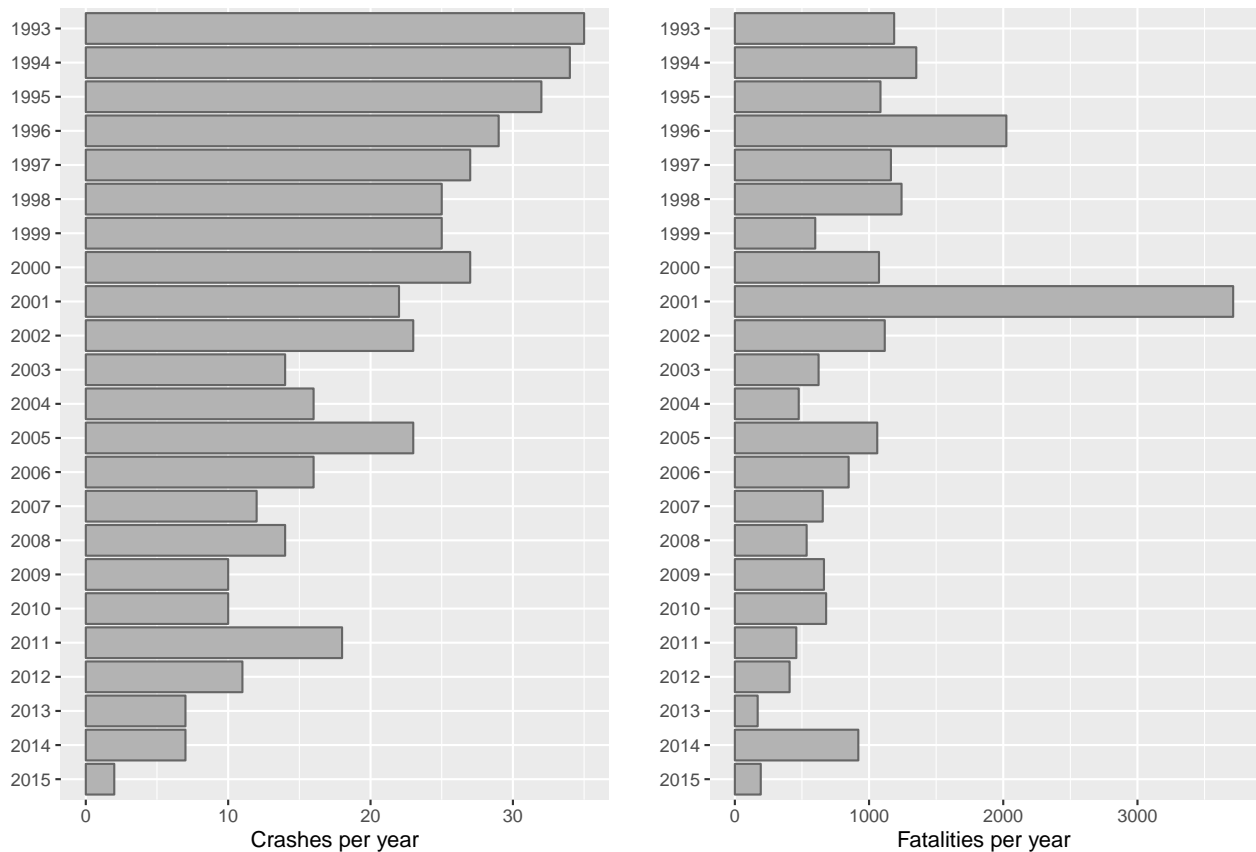


Fig 1: The number of fatal crashes has been declining over the last 20 years, the number of fatalities less so. The worst year for fatalities was, of course, 2001, when 9/11 occurred.

The causes of the 439 accidents and the flight phases in which they happened are shown in the following barcharts. The orderings of the categories for the two variables would be alphabetic by default and so had to be specifically defined. Other orderings would be possible, for instance making “criminal” the final category amongst causes or “unknown” the final category amongst phases.

```
AirCrash <- AirCrash %>% mutate(cause = factor(Cause,
  levels = c("human error", "weather", "mechanical", "criminal", "unknown")))
AirCrash <- AirCrash %>% mutate(phase = factor(Phase,
  levels = c("unknown", "standing", "take-off", "en route", "landing")))
b1 <- ggplot(AirCrash, aes(cause)) + geom_bar() + ylab("")
b2 <- ggplot(AirCrash, aes(phase)) + geom_bar() + ylab("")
grid.arrange(b1, b2, ncol=2)
```

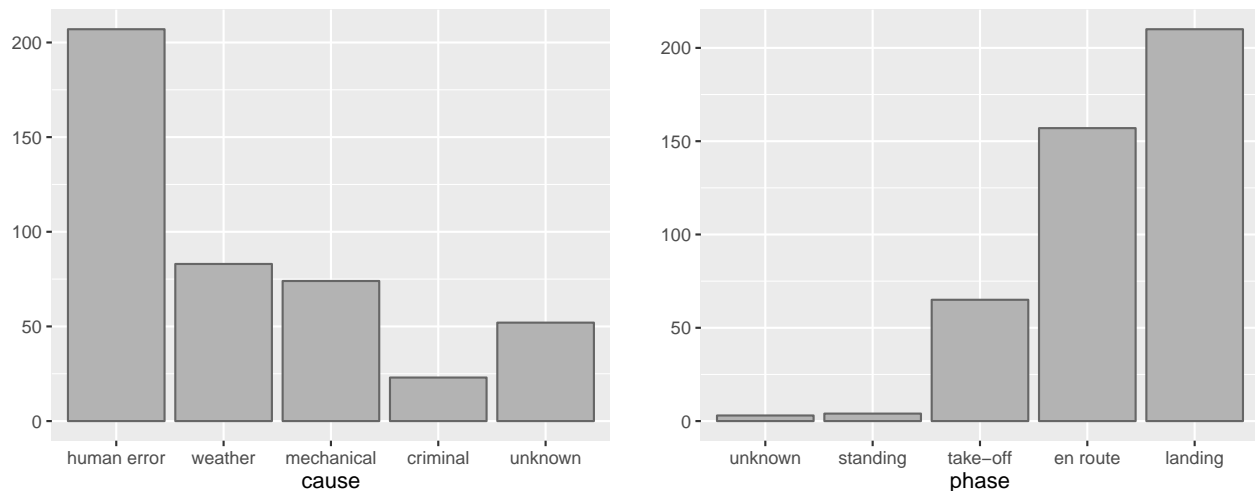


Fig 2: The main cause for crashes was human error. It is a little surprising that the cause of some 12% of the crashes is recorded as unknown. Most of the crashes took place on landing or en route.

The main causes of crashes are different at different stages of flights, as can be seen in the following multiple barcharts display.

```
ggplot(AirCrash, aes(cause)) + geom_bar() + facet_grid(phase ~.) +
  theme(legend.position="none") + ylab("Phase")
```

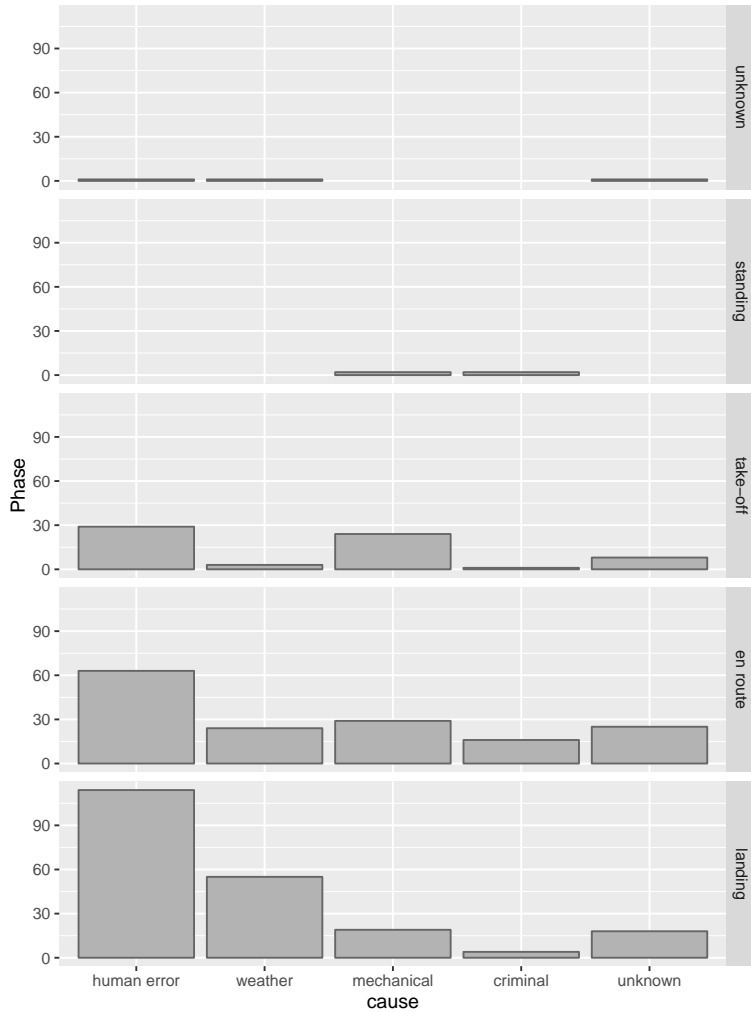


Fig 3: Barcharts of the crash causes by the flight phases. Human error is the main cause for the three main flight phases (take-off, en route, landing). Weather is a problem en route and for landing, but less so at take-off, presumably because planes do not take off in very bad weather. Mechanical faults are relatively less of a problem when landing, perhaps surprisingly. Criminal actions are more of a problem en route than in other phases.

In Figure 1 it was clear that the numbers of crashes with fatalities have declined over the years. Has this been due to declines in any particular cause?

```
ggplot(AirCrash, aes(year)) + geom_bar() + facet_grid(cause ~.) +
  theme(legend.position="none") + ylab("Cause")
```

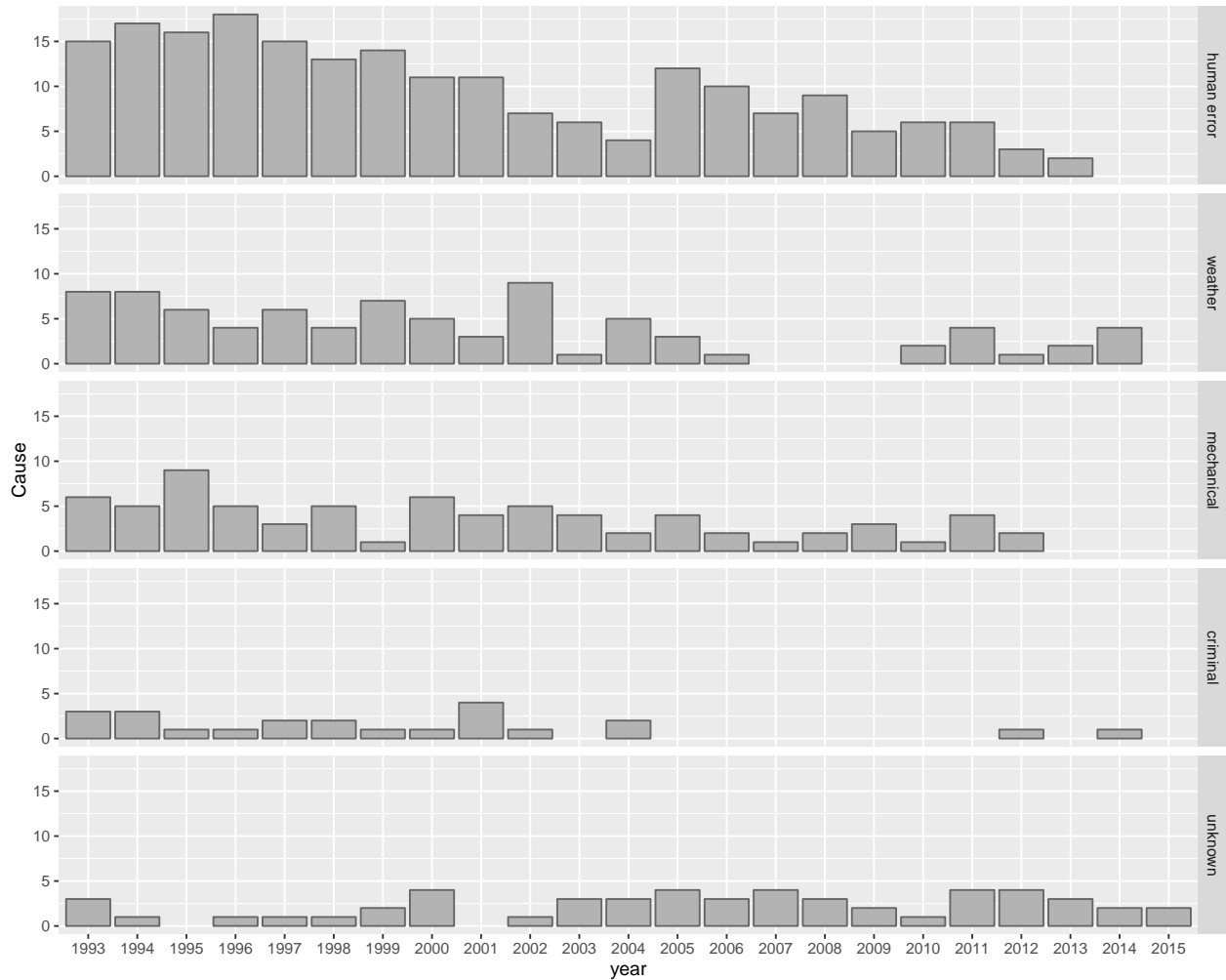


Fig 4: Barcharts of the numbers of crashes by cause per year. The number of crashes due to human error has declined, as have the numbers due to other causes. To some extent this may be due to more causes categorised as unknown. Could this be due to investigations not yet being fully completed?

There are alternative plots for this dataset on its help page in the R package **vcdExtra** and there is an excellent discussion on Rick Wicklin's blog (<http://blogs.sas.com/content/iml/2015/03/30/visualizing-airline-crashes.html>).