

Poor harvest in India in 2002 shown with wheather Data Science

Tilman Hinnerichs, Anja Reusch

Bertelsmann Data Science Scholarship Program

July 21, 2018

What question would we like to answer?

On the search for good data we stumbled accross the CDC (Climate Data Center) of the DWD (German Wheather Center). They are providing data from all kinds of different wheather stations all around the world as we visualized in Figure 1.



Figure 1: Wheather stations provided by the CDC

What question would we like to answer?

The CDC is providing all different kinds of data for those stations. Their database holds the statistics on snowfall, precipitation and sunshine hours and literally everything. The data is provided in a raw heap of data which we tried to classify.

But what to do with that giant pile of data?

Hence, we found another source of data: The **Food and Agriculture Organization of the United Nation**, which provides data on quantitative measurements of the yield of harvest. There we found a peak in the yield of paddled Rice of India in 2002, seen in Figure 2, where more than 22% of the production got lost.

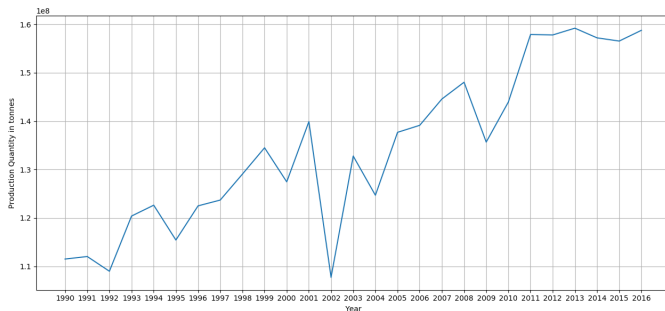


Figure 2: We can see a harsh peak at 2002

What question would we like to answer?

So we would like to know:

- ▶ Can we see possible reasons for this peak in our wheather data?