Ideal approach

- Subhada Suresh sent me this note:
  - (my comment)This was what I'ld like data science to be like when people take their time to do it.
- John Tukey called this "peeling the onion"

Looking at hurricane data

- Stage 1: Thinking as a statistician
  - What is the underlying distribution?
  - Hurricane count by decade

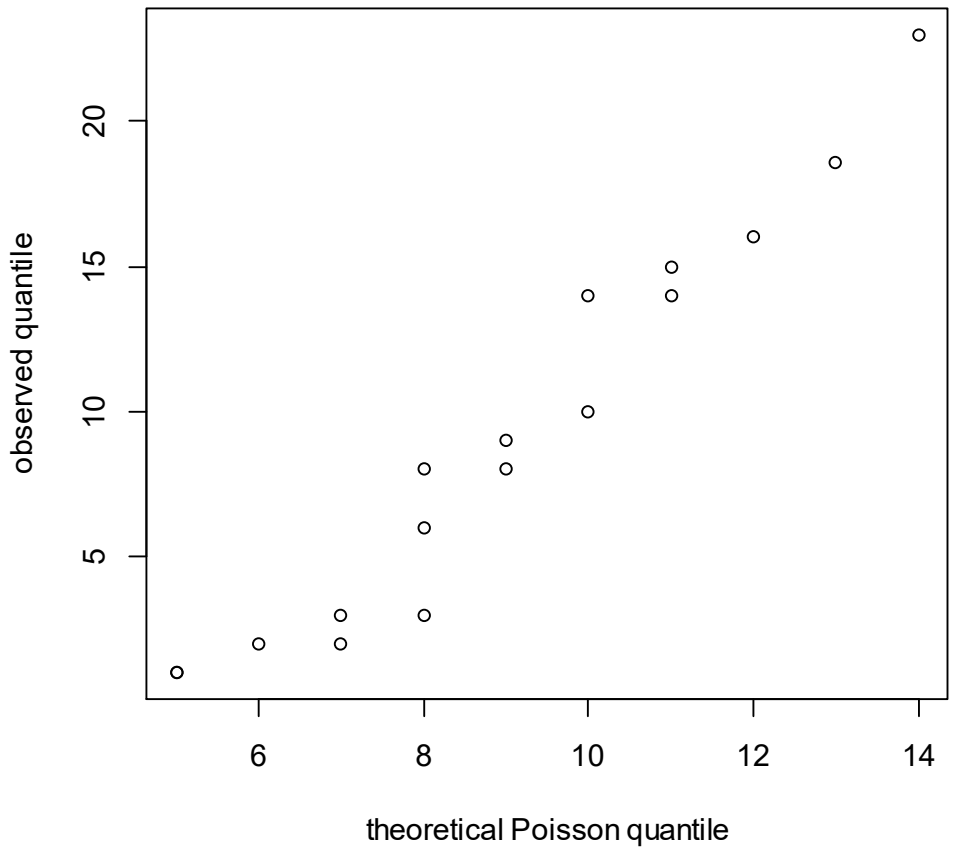
Year	total Count	iv	v
1850	2	2	
1860	1	1	
1870	1	1	
1880	4	4	
1890	4	4	
1900	2	2	
1910	6	6	
1920	8	6	2
1930	16	10	6
1940	9	9	0
1950	15	13	2
1960	14	10	4
1970	8	5	3
1980	10	7	3
1990	14	12	2
2000	23	15	8
2010	13+	11	2

Testing goodness of fit

- Ignore time
- Use KS test (not many data points)
- Use histogram vs theoretical PMF (not best)
- Use QQ plot,
  - Use pencil rule, not rigorous but works
- What are quantiles (on board).

Distribution ignoring time

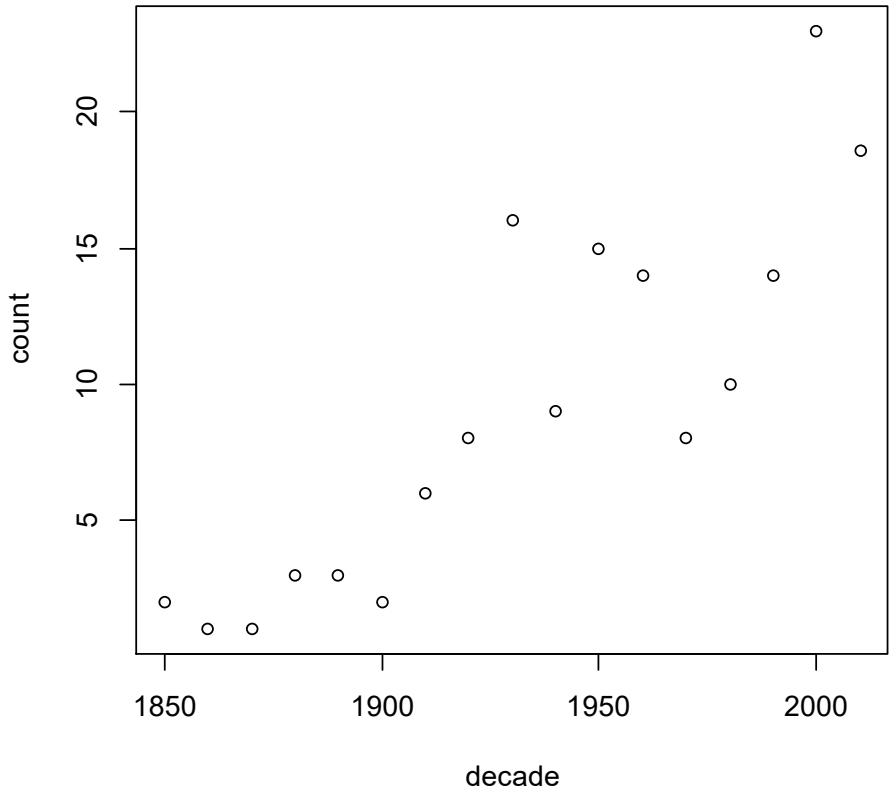
```
> v1<-
      c(2,1,1,3,3,2,6,8,16,9,15,14,8,10,14,
      23,13*10/7)
> v1
 [1]  2.00000  1.00000  1.00000  3.00000
      3.00000  2.00000  6.00000  8.00000
 [9] 16.00000  9.00000 15.00000
      14.00000  8.00000 10.00000
      14.00000 23.00000
[17] 18.57143
> sort(v1)
 [1]  1.00000  1.00000  2.00000  2.00000
      3.00000  3.00000  6.00000  8.00000
 [9]  8.00000  9.00000 10.00000
      14.00000 14.00000 15.00000
      16.00000 18.57143
[17] 23.00000
> mean(v1)
[1] 9.033613
> p1<-c(1:17)/18
> qpois(p1,9.033613)
 [1]  5  5  6  7  7  8  8  8  9  9 10 10 11 11
      12 13 14
> plot(qpois(p1,9.033613),sort(v1))
>
      plot(qpois(p1,9.033613),sort(v1),xla
      b="theoretical Poisson quantile",
      ylab="observed quantile")
>
```



Ignoring time appears to fit poisson
But are counts stable over time

```
plot(10*c(185:201
),v1,xlab="decade
",ylab="count")
```

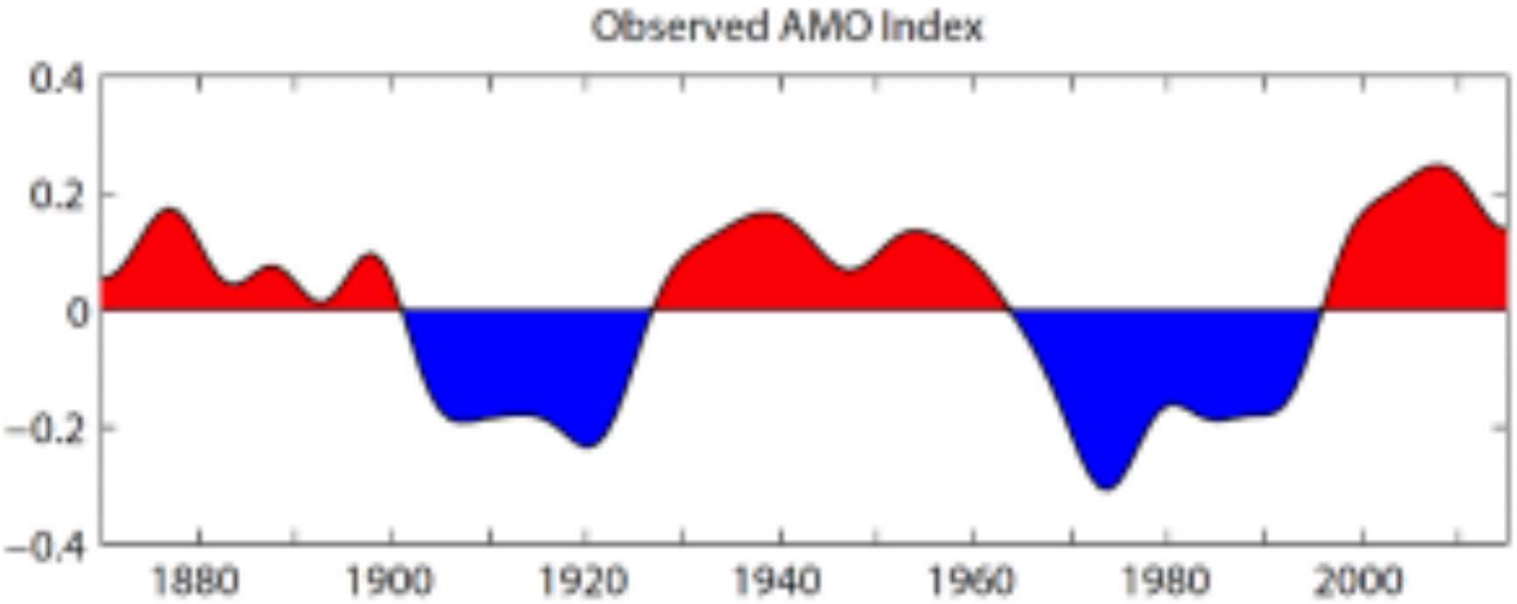
This could be a result of observing technology



What we just saw

- Increasing the set of what we can condition over can change our understanding of the data set
- In the extreme this becomes “Simpsons Paradox”
- Unconditional, strong statistical test in one direction
- Conditionaly for each conditioning variable, strong statistical test in the other direction.

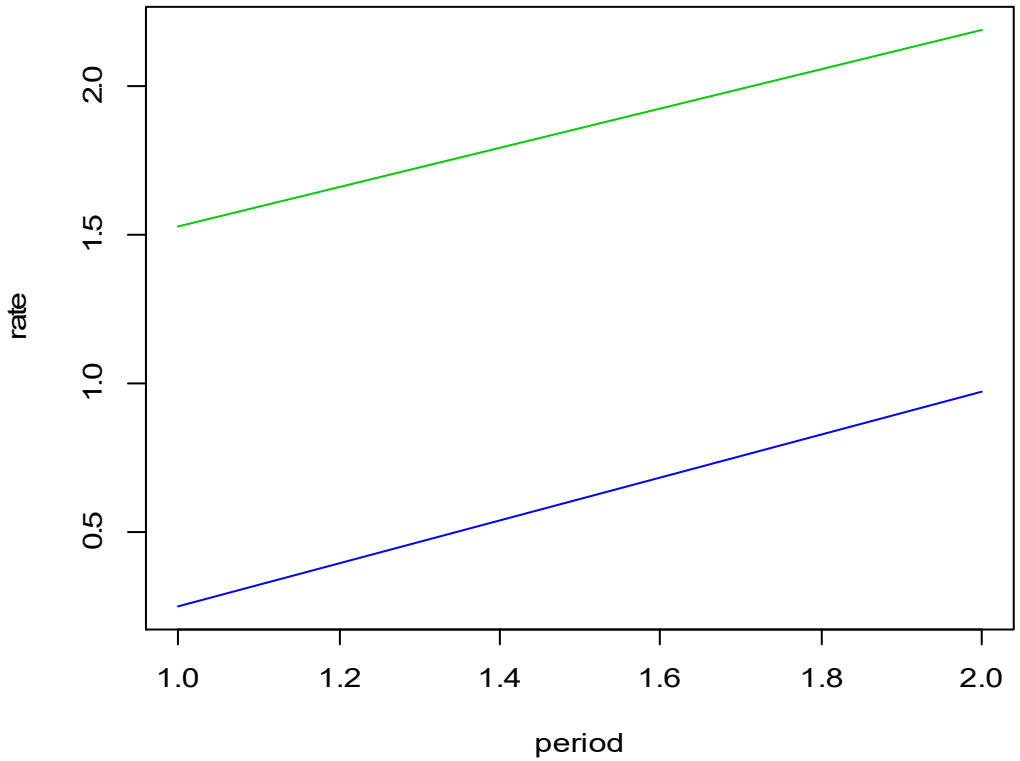
AMO


Observed AMO Index
0.4
0.2
0
−0.2
−0.4
1880
1900
1920
1940
1960
1980
2000

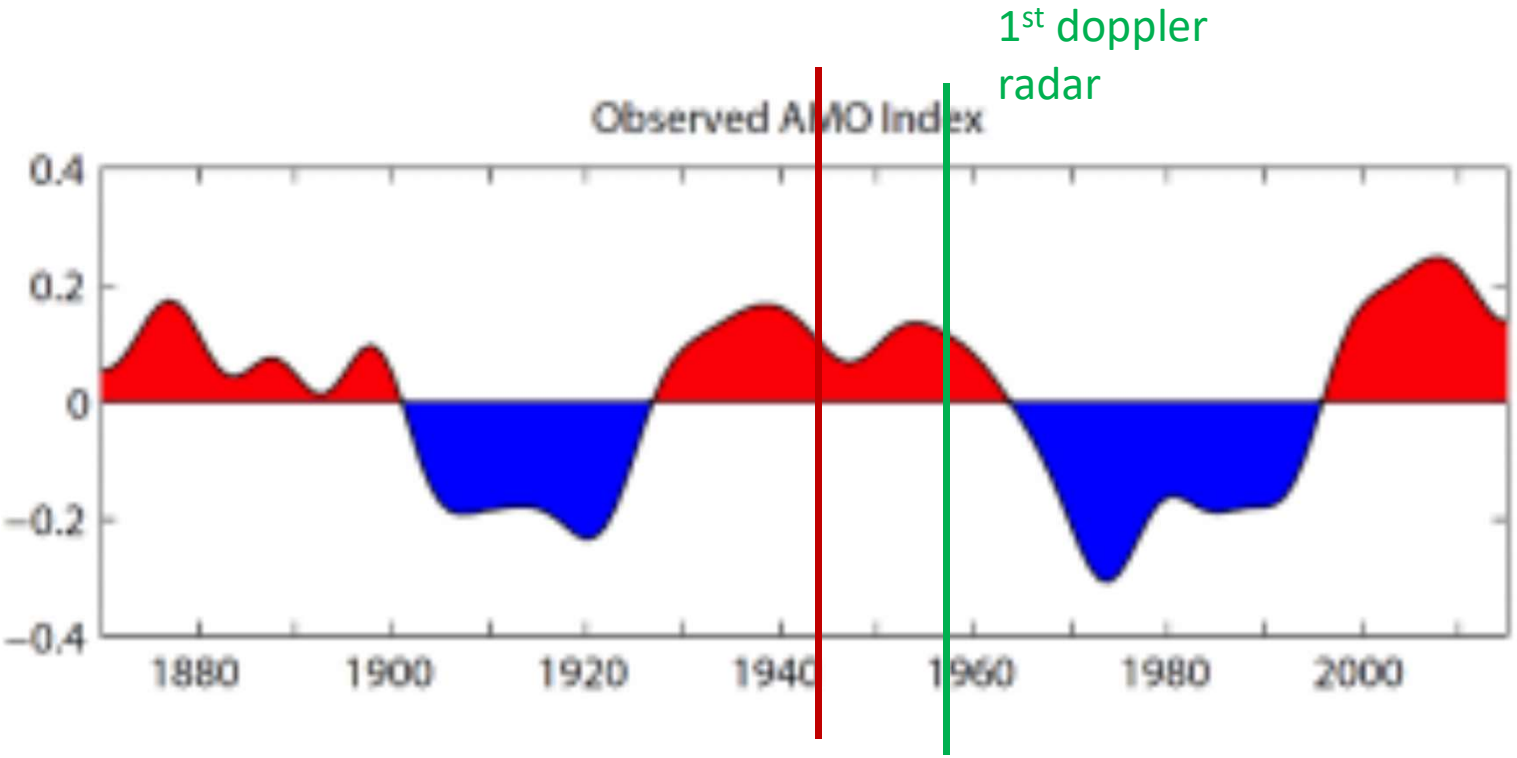
Rough categories

- Cold 1: 1900-1925 V:1 IV 8 years (26) 9/26
- Warm 1 1926-1962 V:11 IV 41 years(36) 55/36
- Cold 2:1963-1996 V:9 IV 23 years (33) 32/33
- Warm 2: 1996- Present V:12 IV 34 years(21) 46/21

Green (warm cycle) blue (cold cycle)



AMO vs Wind speed technology



Conclusions

- If we happened to catch all the hurricanes forming in the atlantic basin, there is evidence of increasing high energy hurricanes. But we've only had one actual AMO cycle with technology good enough to catch most hurricanes.
- Plausible that hurricane strength is increasing but not proven!