

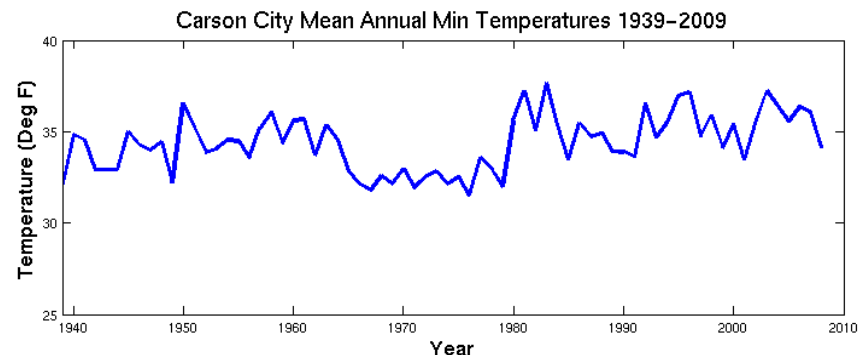
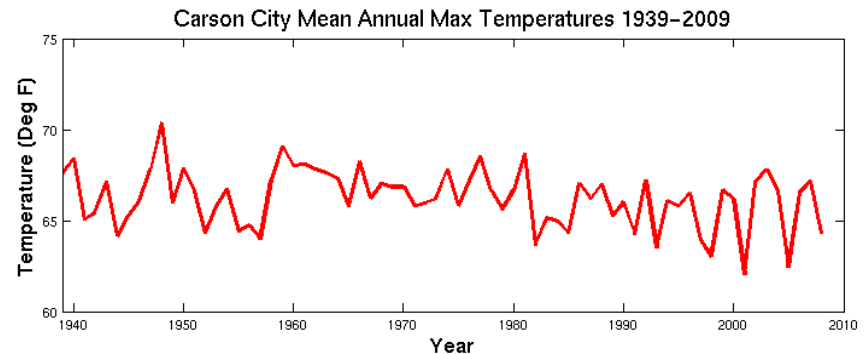
# Historical Climate Data: The Never-Ending Battle For Acquisition

Ben Hatchett, John Mejia, and Darko Koracin

Division of Atmospheric Sciences, DRI, Reno NV

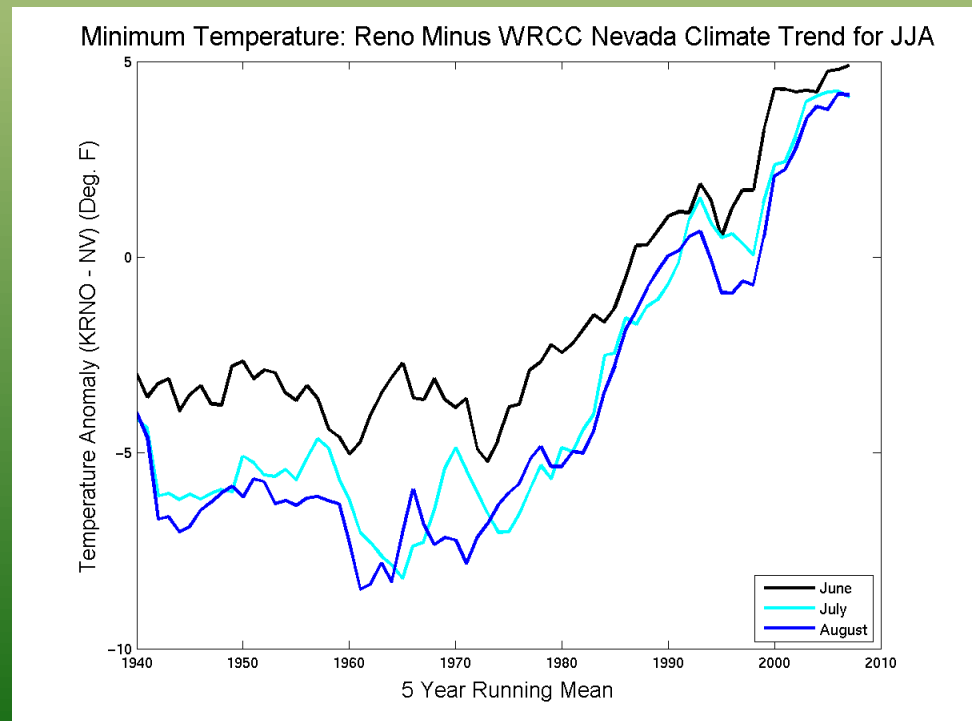


7.17895 268.09633802000 1 12 18 5.648571 -5.167779 268.26198702000 1 13 0 6.704092 -4.183635 269.2139610  
255.53435602000 1 18 12 11.299592 2.937731 256.30756302000 1 18 18 11.377889 1.502851 261.35118302000 1  
171502000 1 24 6 1.930052 -4.843189 259.96967202000 1 24 12 2.761504 -3.151710 255.80068602000 1 24 18 0  
1 30 0 5.501153 -1.979633 271.77785502000 1 30 6 6.057486 -3.599087 268.10359002000 1 30 12 4.541455 -2  
14 2.218626 262.07908502000 2 5 0 0.208735 -0.571388 266.50054602000 2 5 6 -2.064468 -0.480351 252.311  
90 0.706321 243.54466002000 2 10 18 -8.235760 2.139717 251.68339902000 2 11 0 -9.844300 -2.292488 256.248  
5 255.53865502000 2 16 12 2.187994 -5.043405 252.65717902000 2 16 18 2.835452 -2.164085 259.89448402000 2  
6 0.672821 -4.867051 255.67366202000 2 22 12 1.575836 -3.413975 250.92844202000 2 22 18 -0.114206 -1.20918  
273.13058102000 2 28 6 -3.289739 7.495184 270.21606202000 2 28 12 -3.105457 9.622856 269.19206802000 2  
04902000 3 6 0 -1.951059 7.753770 271.65636602000 3 6 6 -1.143320 10.459102 261.37560202000 3 6 12 0  
06712 -9.307067 270.28899702000 3 12 0 -0.654080 -8.422177 270.61234402000 3 12 6 0.716247 -7.359509 25  
8.636411 265.91830802000 3 17 18 -1.007145 7.138856 274.21919602000 3 18 0 -1.808815 8.275999 273.703351  
3 23 12 -3.277577 5.957805 273.81464302000 3 23 18 0.496923 5.674726 274.58795002000 3 24 0 3.849860 2.  
2 -4.903596 259.47177502000 3 29 12 4.161933 -4.104178 256.82488902000 3 29 18 2.451154 -0.544019 271.3586  
1602000 4 4 6 -4.729384 0.142920 266.59600802000 4 4 12 -5.583424 4.038986 265.58102802000 4 4 18 -3.  
1.092034 1.704402 282.00764702000 4 10 6 -6.970341 3.784823 274.84995202000 4 10 12 -6.099239 5.545379 27  
9665702000 4 16 6 6.817656 -0.634910 274.44371302000 4 16 6 6.072042 -2.242727 273.84136402000 4 16 12  
18 -4.314257 5.058750 288.08444102000 4 22 0 -1.743337 6.043232 289.17011102000 4 22 6 6.430754 3.60753  
93 277.733543602000 4 27 18 -2.176019 3.863564 283.33695802000 4 28 0 -2.681200 3.468127 285.24143302000  
02000 5 3 12 -5.395196 -3.790096 278.24165202000 5 3 18 -4.797420 -0.637135 284.69755202000 5 4 0 -4.  
6 0.535356 -2.944547 279.49284602000 5 9 12 -1.170278 0.239380 277.23541102000 5 9 18 -2.833381 2.1796  
280.91055602000 5 15 6 -0.732487 -2.907825 275.30878002000 5 15 12 -1.026049 -3.328493 273.16093002000  
9962102000 5 21 0 0.786383 -10.496590 288.75930402000 5 21 6 -2.498994 -7.865008 276.08067702000 5 21 1  
66526102000 5 26 18 -1.399333 -3.595631 298.08511202000 5 27 0 0.414913 -3.980356 302.25507902000 5 27  
977502000 6 1 12 -0.606462 0.180672 280.26313302000 6 1 18 0.148912 -1.714694 302.38627102000 6 2 0  
1.608209 0.649707 285.36870202000 6 7 12 -1.456170 1.391547 281.22903402000 6 7 18 -2.099967 2.734309 30  
-2.989065 296.78023202000 6 13 6 -3.559576 -0.602113 284.40243702000 6 13 12 -4.301984 5.055927 281.804  
2817002000 6 19 0 6.734100 -1.762525 296.13458002000 6 19 6 3.204172 -2.629197 284.04058302000 6 19 12 1  
03702000 6 24 18 -2.606797 1.047525 292.44051002000 6 25 3 -3.004022 -1.319339 295.65869502000 6 25 6  
30 12 -1.899537 -8.498880 286.07889902000 6 30 18 2.426863 -5.376975 302.84470202000 7 1 0 7.172507 -5.  
94 -4.679727 292.66946402000 7 6 12 -3.170444 -0.812168 286.61981202000 7 6 18 3.097646 0.298637 310.328  
1581 311.71078002000 7 12 6 -8.190931 0.238672 294.27357602000 7 12 12 -4.975322 6.380970 290.0008400200  
05802000 7 18 0 9.636760 -2.098708 310.12252402000 7 18 6 3.981554 -12.435129 292.61872502000 7 18 12  
23 18 6.804247 -2.425910 303.71476402000 7 24 0 6.074159 -3.760070 306.08696102000 7 24 6 -1.163530 -7.  
12 -8.124202 -3.112862 288.16961202000 7 29 18 -4.021154 0.384041 307.02356102000 7 30 0 2.616643 0.9883  
09 -9.247030 293.85729302000 8 4 12 -2.650981 -6.482701 288.06172202000 8 4 18 2.465464 -4.152480 306.3  
71 311.26429502000 8 10 6 7.103558 2.445665 292.58346102000 8 10 12 5.057389 -2.607036 287.10516502000 8  
8 16 0 4.569787 -1.586701 308.53782402000 8 16 6 -3.913005 -6.773363 290.61889802000 8 16 12 -5.414338  
5 4.377492 304.64539602000 8 22 0 8.512585 2.330232 307.46914702000 8 22 6 3.134890 -5.942513 289.7749  
25 283.88490202000 8 27 18 -2.426209 -0.247311 304.73358002000 8 28 0 4.247950 -1.421011 308.03614102000  
2 5.484929 -0.306452 284.44870702000 9 2 18 7.865716 1.724438 298.26570802000 9 3 0 7.927511 -3.072218  
1.266827 291.17489902000 9 8 12 4.261751 8.750519 285.52453802000 9 8 18 8.164381 9.345571 299.43505602  
57 305.97555302000 9 14 6 -0.492268 3.167983 286.25260002000 9 14 12 3.487795 5.440055 283.94144902000 9  
0 8.363303 -3.457534 293.75165302000 9 20 6 -2.140039 -10.821768 281.11821302000 9 20 12 -2.149653 -10.3  
87664 8.604670 281.46880902000 9 26 0 7.762130 8.049372 287.05166402000 9 26 6 6.885576 0.389289 281.032  
18 275.65868602000 10 1 18 3.034639 1.702055 289.82289802000 10 2 0 4.061883 0.961712 294.58257002000 10  
0 10 7 12 4.713866 -6.656109 276.00812602000 10 7 18 5.464453 -5.843277 276.64835902000 10 8 0 6.052889  
5 2.365119 -4.154253 272.48302602000 10 13 12 -0.110756 -0.490982 270.45958402000 10 13 18 1.385533 1.018621  
3 283.40002802000 10 19 6 3.817615 0.104460 277.31215102000 10 19 12 1.951228 3.253370 276.76958102000 10 1  
10 25 0 9.639048 0.732059 285.18241802000 10 25 6 8.136474 4.675700 278.96637102000 10 25 12 11.884246 5.  
33793 272.95662502000 10 31 0 9.401733 -1.558919 272.84477202000 10 31 6 5.965927 -5.915172 270.1718960200  
02000 11 5 18 5.540007 7.889204 280.08545502000 11 6 0 8.700502 2.13116 284.29407502000 11 6 6 5.92024



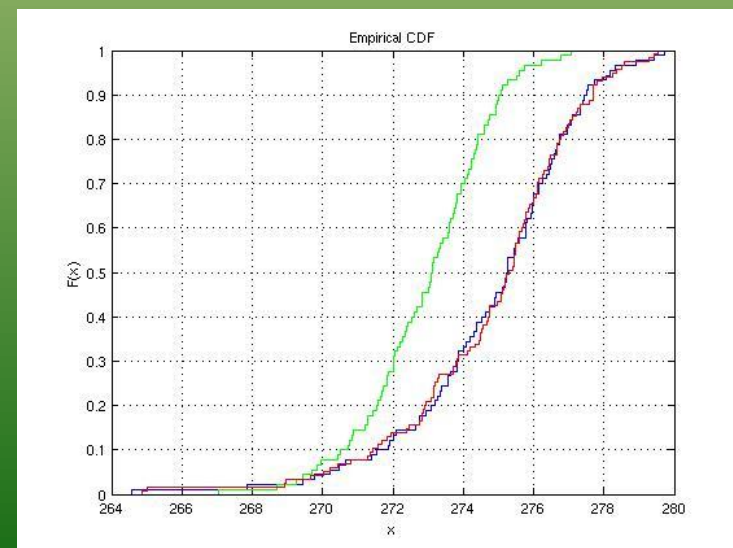
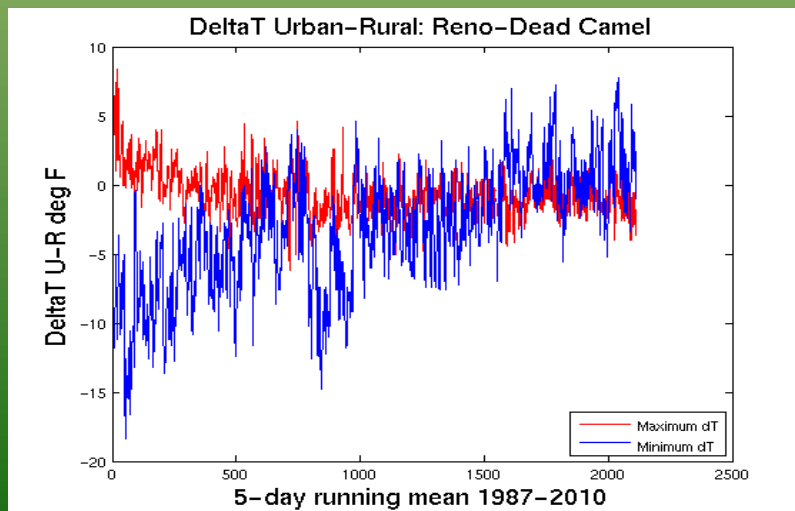
# Outline

- Why do we need historical climate data?
- Where to acquire data?
- Data Issues
- Formatting For Everyone
- Conclusions
- Continuing Work



# Why do we need historical data?

- Hard to understand present or future without knowledge of the past
- Calibration of Models (Downscaling Products)
  - E.g. CDF technique
- Verification of Model Hindcasts
  - Force with even years, verify against observed odd years
- Who gets to deal with acquiring the data?
  - Usually graduate students



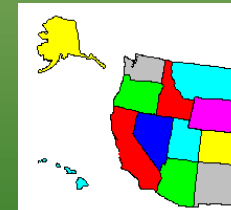
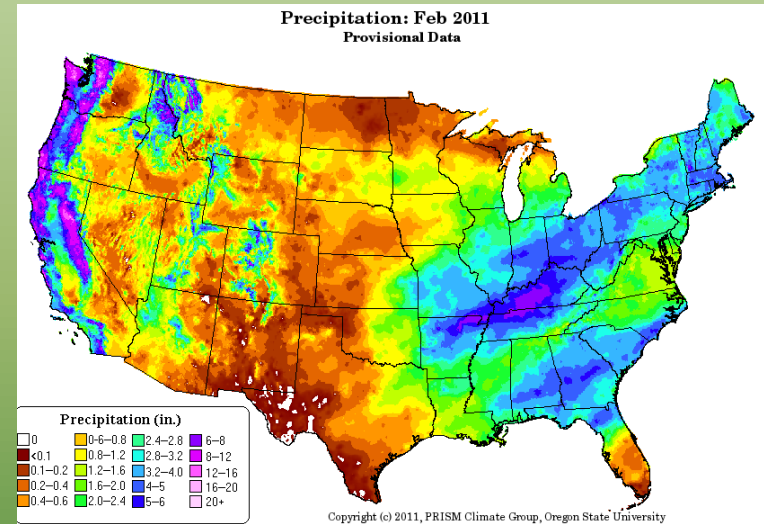
# Where to Acquire Data?

## Obvious Questions To Guide Your Search:

- What temporal scale needed?
  - 10-Minutes? Hourly? Daily? Monthly? Seasonal?
- What spatial scale(s)?
  - Station (point)-based
    - Single station or interpolation?
  - Gridded
    - Further downscaling necessary?

## – EXAMPLE DATA SOURCES

- Western Regional Climate Center (stations)
  - <http://www.wrcc.dri.edu/index.html>
- National Climatic Data Center (both)
  - <http://lwf.ncdc.noaa.gov/oa/climate/climatedata.htm>
- PRISM Climate Group (gridded)
  - <http://www.prism.oregonstate.edu/>
- Earth System Research Laboratory (gridded)
  - <http://www.esrl.noaa.gov/psd/data/gridded/>



Western Regional  
Climate Center



Earth System Research Laboratory  
Physical Sciences Division

NOAA Satellite and Information Service  
National Environmental Satellite, Data, and Information Service (NESDIS)

National Climatic Data Center  
U.S. Department of Commerce

Updates to most datasets on our ftp site will be delayed until later this week.

DOC > NOAA > NESDIS > NCDC

Search Field:  Search NCDC

Current Events  
NOAA Climate Services  
About NCDC  
In the Spotlight

Online Climate Data Directory

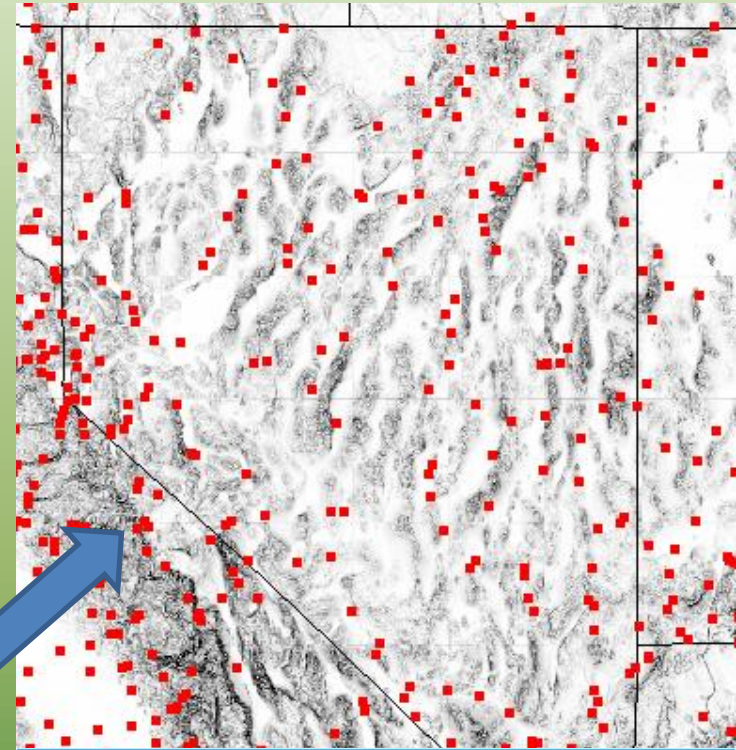
# Data Issues

- *Simply downloading and using data is recipe for trouble.*
  - Understanding problems and learning to recognize them can be useful tool in all disciplines.
- Types of Problems:
  - Erroneous Data
    - Example: Carson City, Nevada  
Maximum T missing, but MaxT reported in MinT column! Solutions often simple yet creative.
  - Missing Data
    - What proportion is missing?
      - Weighting stations based on missing data proportion
    - How are missing data represented?
      - More importantly, is it easy to alter to fit your needs?
      - Scripting to eliminate/replace strange representations
    - Replacing Missing Data
- Note to Students: Learn to script. Bash, Perl, Python, whatever, it will make your life easier! It is never too late!



# Some Bigger Problems...

- Station moves
  - Check the metadata!
- Local Land Use Change
  - Photographs, site visits, land use maps
- Station Locations
  - Automated sensors vs. COOP stations
    - Remote vs. Populated: Topographical Bias
    - Length of Record
      - Most mesonet/RAWs stations >1980
      - COOPs <1890
    - Temporal
      - COOP: daily, automated: hourly/10-minute
    - Quality of data
      - Tipping buckets freeze → Precipitation Bias
      - Anemometers icing up
      - RH <100%
      - Human inconsistencies → measurement time

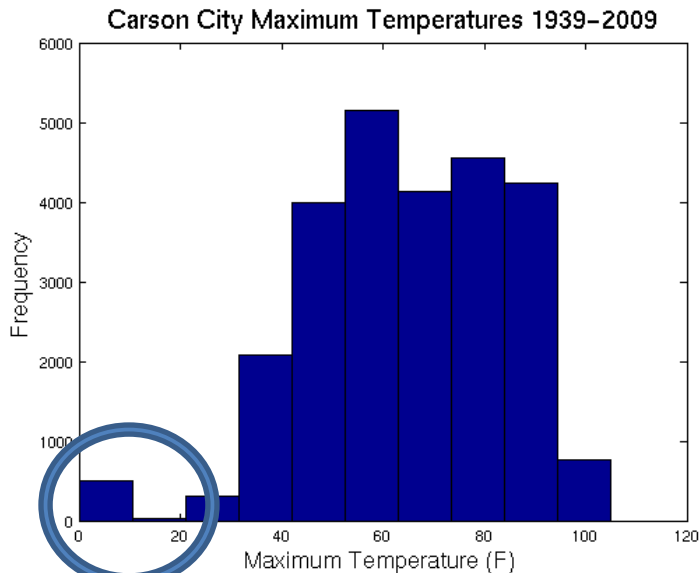
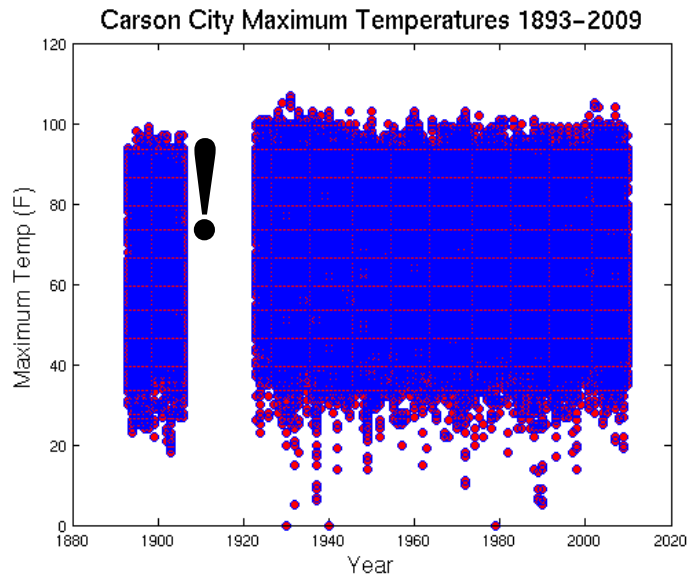


***NV EPScOR transects to the rescue!***



# Tips For Quality Checking

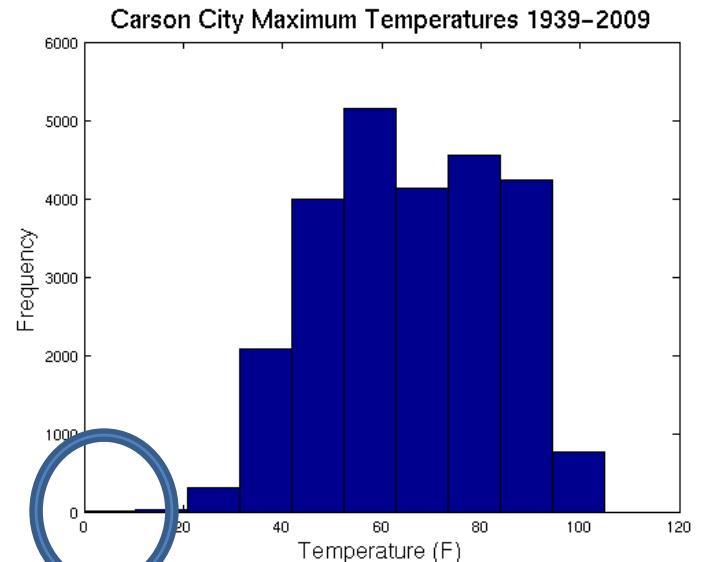
- At each step, make a plot!
- Missing data?
- Strange values/outliers → More investigation!
  - Find the outliers, compare with nearby stations.
  - 0F in Carson, 33F in Reno?
- *Simple, creative solutions will be much less troublesome than over-coding*
  - Helps to count errors as you go
- NCDC tries to flag errors; occasionally miss.
- Up to the users to report strange data!



<3 $\sigma$  DTR → NaN  
40 values found



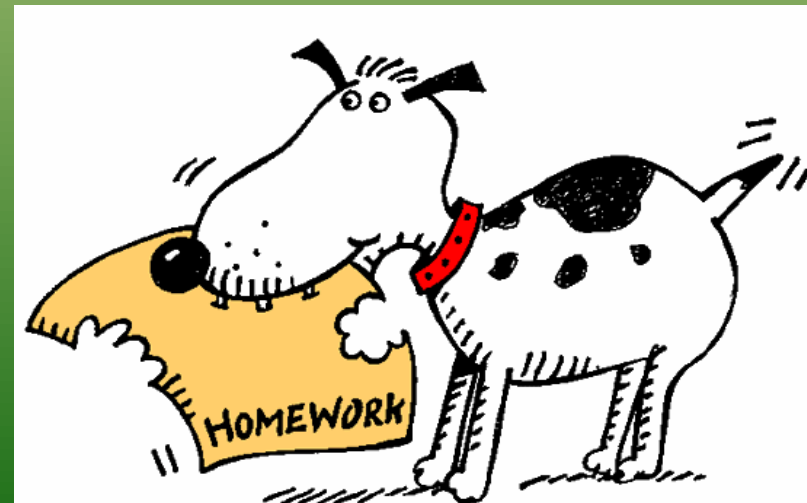
2 errors remain



# Filling in the Blanks...

- Many methods to replace missing data
  - Threshold?
- Simplest method: Replace with mean values
  - Changes distribution
- Regression method (Smith et al. 1996)
- Expectation Maximization (Schneider 2000)
  - Yields better results than non-iterative regression method

<http://www.gps.caltech.edu/~tapio/imputation/>



<http://www.math.hmc.edu/~kindred/math104/graphics/dog-hmwk.gif>



# Formatting For Everyone!

- This is somewhat of a fantasy of science; hard to please everyone
- Main uses: Model Input, GIS, Timeseries Analysis
- Ideas to Reduce Downstream Formatting Issues
  - Discuss with anyone you might work with in future what format they would prefer.
  - Format data initially in manner that can easily be manipulated/loaded into analysis programs
  - Create Metadata!
    - Mention data problems, methods, explanations, etc
    - The better the metadata, the fewer emails those using your data will pester you with

# Good and Bad Formatting

- **Bad Formats...**

- Use one or (worse) many of these: - / , | ,,
- No metadata, converted units
- MM/DD-YYYY,MaxT|MinT,Precip

- **Good Formats...**

- Can easily be read into analysis/visualization tools
  - No fun to spend time manipulating
- Have well-written metadata (units, locations, scales)
- Use familiar units for user group (WRCC in degF)
- Example (ours, so of course its best)
  - Output as ASCII, also write to NetCDF
  - YYYY MM DD MaxT MinT Prec \n
  - Why is this good?
    - It's Simple!!!

# Conclusions

- Climate data is often neither simple nor well-organized
  - Significant time investments may be necessary to achieve state of data desired.
- Think simple and creative
  - Finding outliers or weird data points
  - Removing/replacing erroneous values
- Investing time into manipulating/formatting is not a waste...
  - The downstream cascade of well-formatted, quality data will save time, reduce headaches, and result in higher quality research
  - Everyone benefits from data sharing!
  - Don't hesitate to ask your users about formats/information they desire (preemptive strike!)

# Present and Continuing Work

- Formatted 45 RAWs and 35 COOPs (finishing ALL Nevada COOPs soon) to simple daily format: YYYY MM DD Tmax Tmin Precip
  - Easy for most impact studies to utilize
- Stations are being statistically downscaled and used in hybrid statistical-dynamical downscaling method developed at DRI
  - Part of Intermountain Climate Change Assessment
    - Historical Analysis/Modeled Future for Western U.S.

# References

- Smith, T. M., R. W. Reynolds, R. E. Livezey, and D. C. Stokes, 1996: Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Climate*, **9**, 1403–1420.
- T. Schneider, 2001: Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *Journal of Climate*, **14**, 853-871.