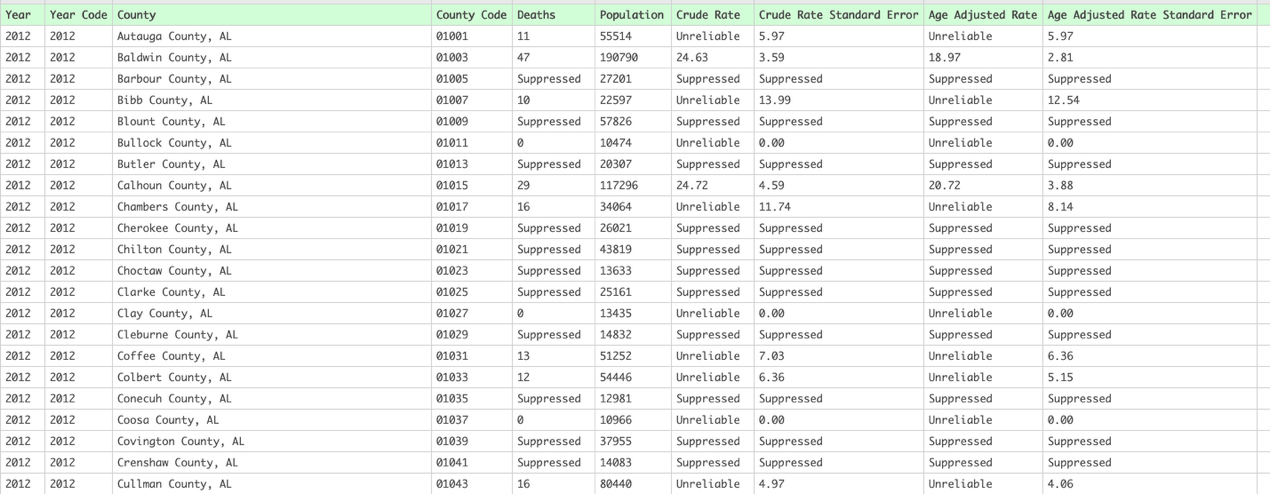
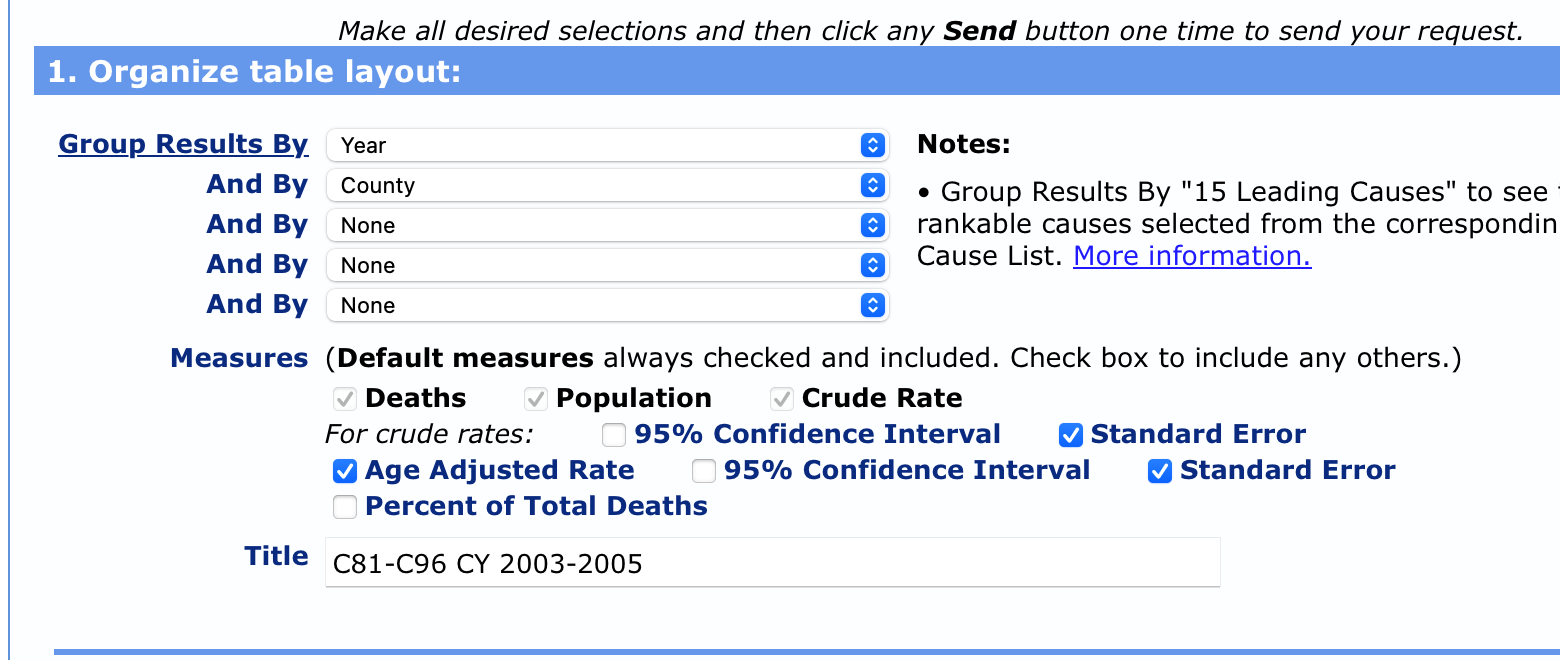
要下两个数据，先说第一个数据：  
数据1: 县×年份的数据：

之前我们下载的是分层的数据：县×年份×性别×种族×年龄，但是由于cdc数据抑制问题，如果某一个层，或者说单元格数值小于10，那就不会显示，比如某县某年白血病死了20个人（13男，7女），那么如果我们下载县×年份的数据就会显示20，但是如果下载县×年份×性别的数据就会显示男的是13，女的就不显示了，然后就得给模型传入这个数值是[0,9]中间的一个值，然后用极大似然估计，就降低了数据的精度。所以对于初始的不分层的分析，我们需要尽可能详细的数据，也就是分类尽量少，所以要下一个县×年份的数据，

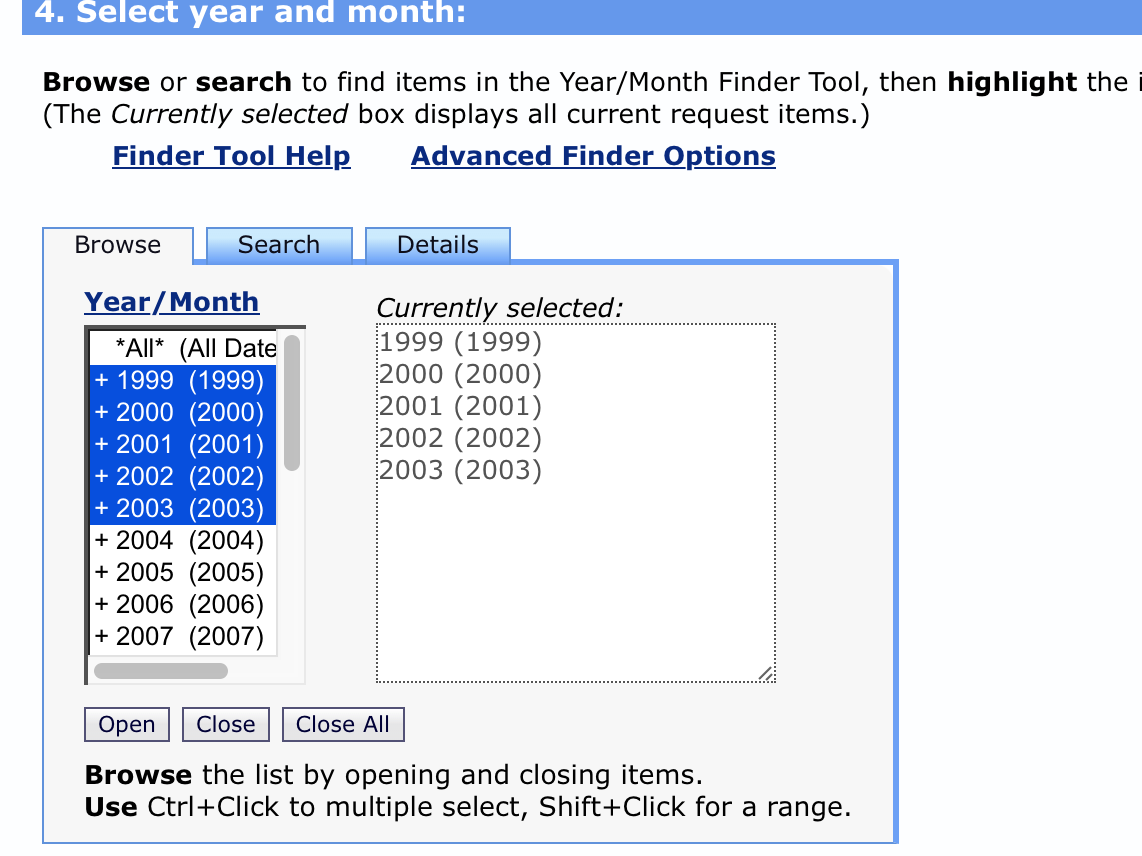
如图，只有县×年份 然后直接就是死亡，这个是算时间序列的分类最少的数据了，下载的时候选：  


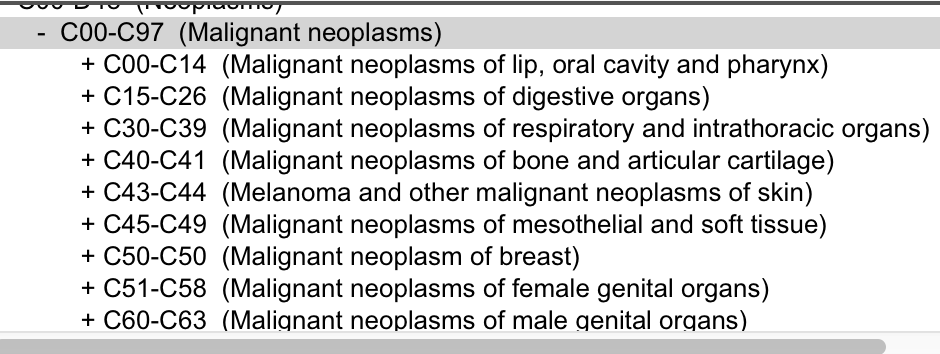
分层选year,county

变量选标准误，年龄调整死亡率，年龄调整死亡率的标准误

命名：疾病编号+CY+年份

CY意思是County Year

这个年份可以选多一点，大的病像肺癌乳腺癌，可能得两三年一选，小的病比如什么非特异性的癌症可以五年一选，看情况来吧，反正超了会提示的。  
 

病的列表还跟之前一样，先把大的都下了吧，然后还有几个小的特定的值得关注的。然后C81-C96不要下了，我已经下了  


单独的病的列表：  
Group1

C18-C21 Colorectal Cancer 结直肠癌(Malignant neoplasm of colon, Malignant neoplasm of rectosigmoid junction, Malignant neoplasm of rectum, Malignant neoplasm of anus and anal canal)

C25 Malignant neoplasms of pancreas 胰腺癌

C34 Malignant neoplasms of bronchus and lung 肺癌和支气管癌（死亡率最高）

C50 Malignant neoplasms of breast 乳腺癌（女性最常见）

C61 Malignant neoplasms of prostate 前列腺癌（男性最常见）

Group2

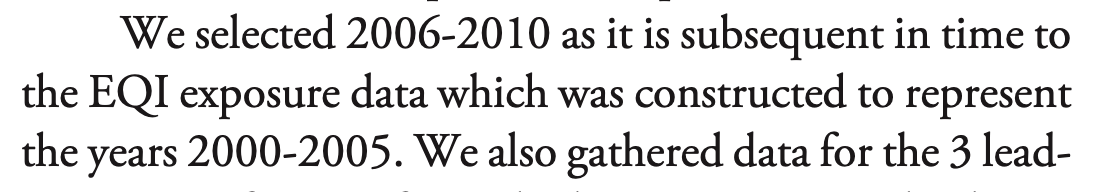
C22 Malignant neoplasms of liver and intrahepatic bile ducts 肝癌和肝内胆管癌

C64 Malignant neoplasms of kidney, except renal pelvis 肾癌，不包括肾盂恶性肿瘤

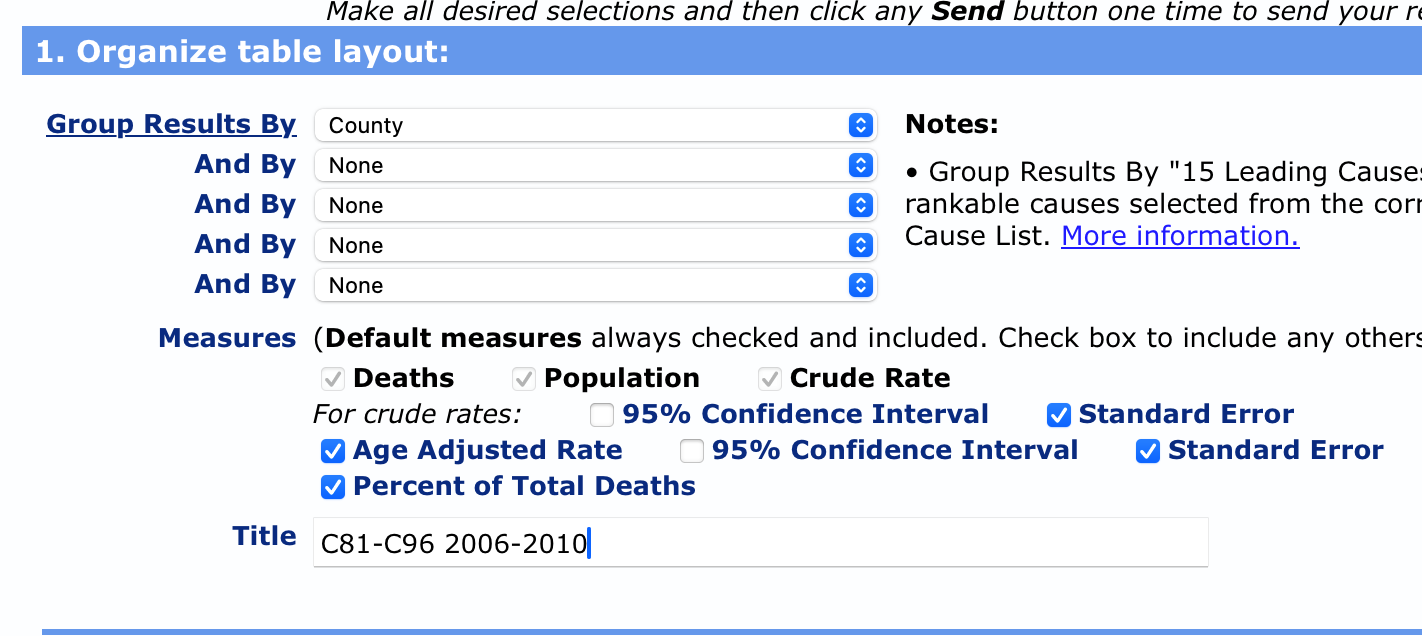
C67 Malignant neoplasms of bladder 膀胱癌

C82-C86 Non-Hodgkin lymphoma 非霍奇金淋巴瘤

C91-C95 leukemia 白血病

第二个数据：五年连续数据  


美国环保局的一个数据 环境质量指数 EQI，这个数据是2000-2005年的6年平均数据，Cancer上面一篇文章用的滞后五年的发病率分析的，我们准备做死亡率，所以分别下五年,十年,十五年滞后，也就是说，我们需要下载2006-2010年和2010-2015年和2015-2020的死亡数据

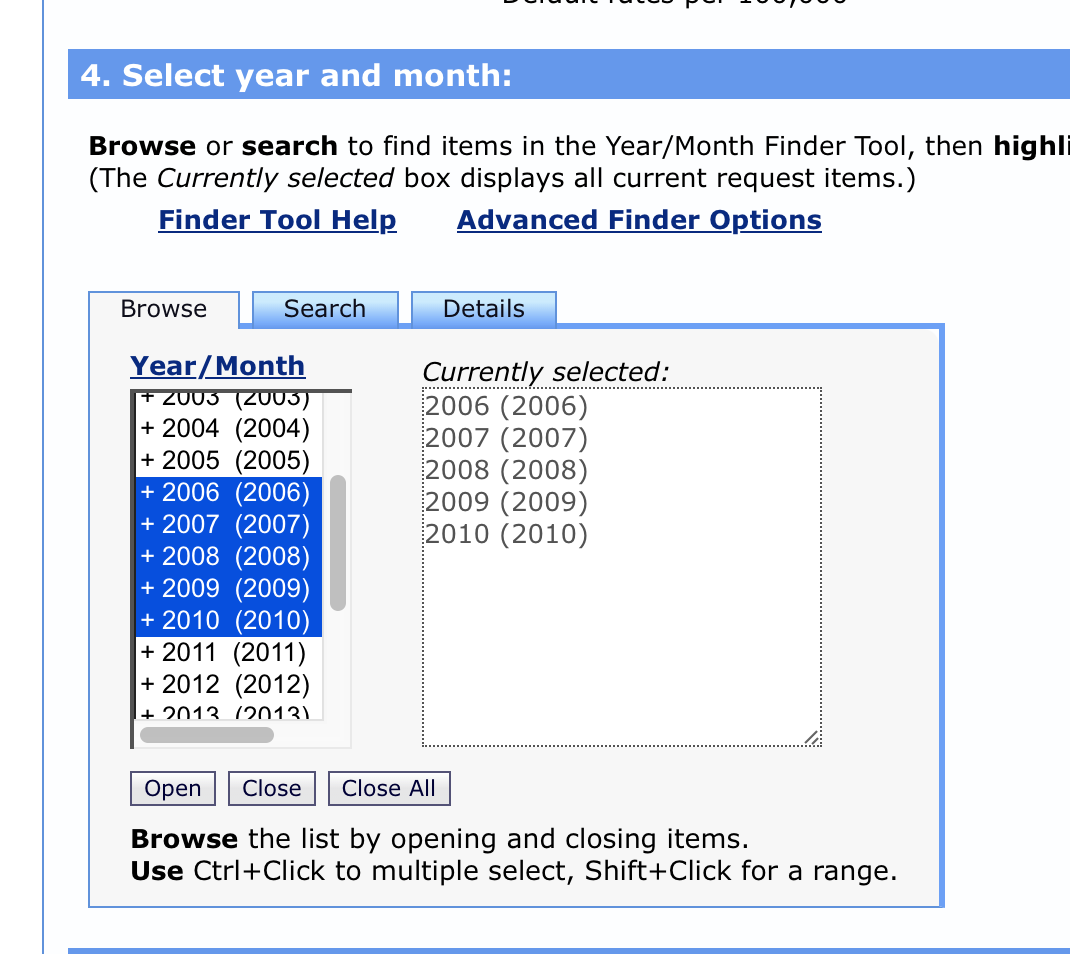


注意这里的**分层不要选year**，因为我们要六年合并的数据，只选county，下面选的一样，变量选标准误，年龄调整死亡率，年龄调整死亡率的标准误

命名： 疾病+2006-2010（5年滞后）

命名： 疾病+2010-2015（10年滞后）

命名： 疾病+2015-2020（15年滞后）



年份就选我们要的年份  
  
疾病一样的，不变，这个要下C81-C96 我没下

总结一下，最后应该是如下数据：  
县×年份的数据 C81-C96 CY （这个每个病会有多个，因为有多个年份）  
滞后五年的数据 C81-C96 2006-2010（这些都是每个病一个）  
滞后十年的数据 C81-C96 2010-2015

滞后十五年的数据 C81-C96 2015-2020