

## The Color of Lynching

Lisa D. Cook  
Department of Economics and  
James Madison College  
Michigan State University  
October 2011  
[lisacook@msu.edu](mailto:lisacook@msu.edu)

©2011

The author wishes to thank Trevon Logan for helpful comments and Mattan Alalouf and Myoungjin Key for able research assistance. Any mistakes are my own.

## Abstract

Data on lynchings have increasingly been employed in economics research. Historical lynching data sets are difficult to use due to measurement error, particularly with respect to the race or ethnicity of the victim. This paper summarizes recent researchers' efforts to correct and extend data related to race and ethnicity of lynching victims, especially with respect to unpacking the "white" category. A systematic examination of the revised data reveals that historical data sets overestimate the share of whites among lynching victims and underestimate the share of Chinese, Hispanic, and Native American victims by nearly 50 percent.

Violence is a regular feature of many societies and economies around the world. In recent years, interest in violence and economic outcomes has risen among economists and other researchers. Miguel, Satyanath, and Sergenti (2004) use an instrumental-variables approach to show that shocks to economic growth in Africa are correlated with civil conflict in Africa between 1981 and 1999. Many others find or assume the opposite relation. Abadie and Gardeazabal (2003) use GDP and stock-performance data to test the effects of domestic terrorism on economic growth in the Basque region. They find that terrorist activities reduce economic performance by 10 percent. The findings in Collins and Margo (2004) suggest that the riots of the 1960's lowered the value of commercial and residential property. Ferguson (2006) examines the relation between the concentration of violence in Central and Eastern Europe, Manchuria, and Korea and economic volatility, among other factors, in the 20<sup>th</sup> century.

Data on lynchings, an extreme form of violence, are increasingly being used to explain a variety of outcomes. For some time, researchers have analyzed the causal relation between labor-market competition between blacks and whites and lynchings, e.g., Raper (1933), Hovland and Sears (1940), Tolnay and Beck (1995). This literature finds an inverse relation between cotton prices, and therefore competition for jobs in agriculture, and black lynchings. The more recent research finds that this relation breaks down beginning in the early 20<sup>th</sup> century. Fryer and Levitt (2007) use lynching data to study the effects of Ku Klux Klan activity. Carden (2006) interprets the presence of lynching as a signal of insecure property rights and an explanation for the relative underdevelopment of the postbellum South. Using data on patent activity,

Cook (2011a) finds a relation between reduced economic activity, as measured by patents, and lynchings, riots, and segregation laws.

But the question of direction of causality between violence and economic outcomes remains unsettled in the empirical literature. While a number of these papers have attempted to use historical lynching data sets for this purpose, it is widely acknowledged that there are shortcomings in these data, including various types of misclassification and misidentification. Building on Cook (2011b), this paper's contribution is to summarize recent efforts to address measurement error in lynching data, especially related to racial classification. While it remains the case that most victims were African American and most lynchings were in the South, the aggregation of all non-blacks into the "white" category masks significant racial and ethnic heterogeneity across the United States and particularly in the Southwest. From an examination of revised data sets that include the state of California, I find that roughly half of "white" lynching victims were in fact Chinese, Hispanic, and Native American. When accounting for mixed-race victims and those of unknown racial and ethnic origin, the share of black victims also declines relative their share in the historical NAACP data set. Coefficients on lynching-related explanatory variables for whites extracted from the historical series would be larger than anticipated for individuals who were actually white. More importantly, due to measurement error, the lynching variable may be correlated with the error term, and OLS estimators would be biased and inconsistent. Future empirical research would benefit from minimization of measurement error with more careful assignment of all victims' characteristics, including race and ethnicity.

## *Historical Lynching Data Sets*

It is worth defining lynchings before proceeding. While the definition of lynching can vary, the National Association for the Advancement of Colored People's (NAACP's) standard is the one that is widely accepted and is used in this paper. It requires that four conditions obtain: (1) there must be evidence that a person was killed; (2) the person must have met death illegally; (3) a group of three or more persons must have participated in the killing (to rule out personal vendettas, etc.); and (4) the group must have acted under the pretext of protecting justice or tradition.<sup>1</sup>

A number of these data sets are historical. Data on lynchings were taken largely from newspapers. The best known historical data sets were created by the *Chicago Tribune*, Tuskegee University, and the NAACP (see Table 1).<sup>2</sup>

The *Chicago Tribune* started collecting and publishing annual data on lynching victims in 1883. The *Tribune* data have served as the basis for a number of well-known works, including those of Ida Wells-Barnett, the anti-lynching advocate and founding member of the NAACP, in which she

---

<sup>1</sup> The NAACP definition, while widely used in the current literature, evolved over time. In NAACP (1919), the definition of a lynching victim was a person "killed by a lynch mob" (p. 7). The fuller definition cited above was adopted nearly three decades later, which suggests that earlier lynchings reported by the NAACP may not have conformed to the later definition.

<sup>2</sup> Much of this discussion on the general comparison of historical and current lynching data sets is borrowed from Cook (2011b).

makes some of the earliest attempts to explain the causes of lynching by examining the empirical record.<sup>3</sup> Tuskegee University began collecting lynching data in 1892. This data set goes from 1882 to 1968.<sup>4</sup> The NAACP started collecting lynching data in 1912. The NAACP data, which are for the years 1889 to 1918, are compiled in *Thirty Years of Lynching in the United States, 1889–1918* and continued keeping records on lynching in annual reports through at least 1955.

The lists of victims in these data sets contain the victim's name, race (black or white), date and location of lynching, and alleged offense. In recent years, these data have been shown to be flawed. Errors identified include victim misidentification, erroneous reporting of the location, and misclassification of lynching, i.e., the standard definition of lynching was not met.<sup>5</sup>

An important problem that has received attention in the literature recently is that the racial classification in the *Tribune*, NAACP, and Tuskegee data series is largely dichotomous, and victims are categorized as either black or white. The fact that non-black victims of Chinese, Hispanic, Italian, Native American, and others of distinct ancestry are identified as "white" is especially problematic. This would be particularly pronounced in Western and Southwestern states where such populations were larger.<sup>6</sup> Carrigan and Webb (2003) find that mobs lynched nearly 600 people of Mexican descent between 1848 and 1928, which would represent almost

---

<sup>3</sup> See for example, Wells-Barnett [1892] (1969).

<sup>4</sup> The Tuskegee data are also contained in the *Historical Statistics of the United States*, i.e., Carter (2006a).

<sup>5</sup> When reviewing a combined list of the *Tribune*, NAACP, and Tuskegee data for southern states, Tolnay and Beck (1995) find that roughly 17 percent of events identified as lynchings do not meet the modern criteria.

<sup>6</sup> Cutler [1905] (1969) separates out "Others," which includes people of Mexican origin, Native Americans, and foreigners in some of the *Tribune* data. While a small number of Jews are known to have been lynched, e.g., Leo Frank in Georgia, religion is not recorded in any national data set.

half of all white lynchings recorded in the Tuskegee data. The motives for lynching such people would be more heterogeneous than if they were in fact white Americans. It is often suggested in the literature that lynchings in the West were “frontier justice” and that such executions would have been legal if the institutions were developed enough to carry them out, e.g., Pfeifer (2004). Further, most white lynchings are considered consistent with “frontier justice,” having occurred following a simulated or real judicial procedure. Nonetheless, Carrigan and Webb (2003) observe that the lynchings of people of Mexican descent were closer in character to lynchings of African Americans, since the executions occurred shortly after victims were accused of crimes rather than following a trial. They also observe points of divergence from African American lynchings, such as use of guns for execution instead of nooses and the composition of crimes allegedly carried out by the victims.

Even in the rare case when a third racial category was allowed, the classification is still imprecise. Cutler [1905] (1969) separates out “Others,” which includes people of Mexican origin, Native Americans, and foreigners in some of the *Tribune* data. Within this catch-all group it is reasonable to expect significant heterogeneity with respect to lynching motives. Foreigners were more likely to be accused of and lynched for strike-breaking than other racial and ethnic groups.<sup>7</sup>

Since detailed data are not available on white lynchings outside the South, there is no means of systematically separating whites from non-whites in the white lynching data. Analysis using

---

<sup>7</sup> Author’s calculations from Project HAL data.

these historical series that involve race or ethnicity and use lynching data on victims reported to be white require extra caution in estimation and in interpretation.

### *Contemporary Data Sets*

Recent efforts have attempted to address shortcomings of the historical lynching data sets. Tolnay and Beck (1995) carefully analyzed the *Tribune*, NAACP, and Tuskegee data for states in the South, i.e., Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, and Tennessee.<sup>8</sup> Newspaper stories were used to evaluate each case in the combined list and to add cases previously unreported in the data. There are 2,805 observations of confirmed lynching victims in the resulting data set.<sup>9</sup>

A number of researchers have extended the task of improving the quality and scope of historical lynching data to other states. In doing so, they have generally made the data on race, ethnicity, and national origin more precise.<sup>10</sup> Pfeifer (2004) reviews and collects lynching data from various regions of the U.S., i.e., on Alaska, California, Iowa, Louisiana, New York, Washington, Wisconsin, and Wyoming. He verifies and updates data using court records, local newspapers, and other primary sources.

---

<sup>8</sup> The Beck and Tolnay 1997 inventory, which is the basis for but released after Tolnay and Beck (1995), was provided to the author.

<sup>9</sup> The Project HAL database is a digitized version of the data underlying Tolnay and Beck (1995).

<sup>10</sup> Bailey, Beck, and Tolnay (2011) use the Beck and Tolnay 1997 inventory to match lynching victims to census data, which would also provide more detailed information on race and ethnicity.



Gonzalez-Day (2006) examines lynchings in California from 1850 to 1935 and offers to date the most extensive corrections to and information on racial and national-origin classifications of lynching victims in the Tuskegee data. For whites of European origin, he identifies country of origin and, if the victim's birthplace is the U.S., he identifies the state of birth of the victim. Foreigners of European descent – Canadians, Australians, and Europeans – constituted 23 percent of whites lynched. Latinos constituted 38 percent of total lynching victims in these data.<sup>11</sup>

Similarly, in Carrigan's and Webb's (2003) effort to identify people of Mexican descent who were lynched in the 19<sup>th</sup> and 20<sup>th</sup> centuries, they find that, of 36 people reported lynched in New Mexico between 1882 and 1968 in the Tuskegee data, 9 of 33 reported as "white" were of Mexican descent, and one was Native American.<sup>12</sup>

### *California Comparison*

To illustrate the differences between historical and contemporary racial composition of lynching data sets, I use data on the state of California from NAACP (1919), Pfeifer (2004), and Gonzalez-Day (2006). Figures 1 to 3 display the racial differences recorded for California in the NAACP historical data set and in the revised data from Pfeifer (2004) and Gonzalez-Day

---

<sup>11</sup> Author's calculations. Among the Europeans, Italians were the largest group of lynching victims.

<sup>12</sup> Gonzalez-Day (2006), p. 38. Using accounts from diplomatic correspondence, English- and Spanish-language newspapers, and other primary sources, in total, they find that 597 people of Mexican descent were lynched in 13 states between 1848 and 1928.

(2006).<sup>13</sup> The NAACP data in Figure 1 report that, between 1889 and 1908, 92 percent of lynching victims were “white,” and eight percent were African American. In contrast, data from Pfeifer (2004) show that, between 1890 and 1908, 42 percent of victims were white, 13 percent were Latino, 21 percent were Native American, eight percent were Chinese, eight percent were mixed race, and four percent were African American. For the years 1850 to 1935, from the Gonzales-Day (2006) data I calculate that 34 percent were of European descent, 38 percent were Latino, 12 percent were Native American, eight percent were Chinese, six percent were of unknown racial and ethnic origin, and two percent were African American. Although the comparison is not exact, the latter two data sets are closer in racial and ethnic composition than they are to the NAACP data.

The new data sets and similar efforts make it possible to credibly incorporate race, ethnicity, and national origin in econometric estimation. However, such a possibility is still on the horizon, if the outcomes involve national data. As is mentioned in Cook (2011b), efforts to revise historical lynching data sets have evolved piecemeal – by city, state, or region. As a result, there is not yet a national revised database of lynching victims.

### *Conclusion and Further Research*

---

<sup>13</sup> While data have been extracted from the Pfeifer data set for a period similar to the NAACP data set, an exact comparison with the Gonzalez-Day data is not possible, given the different time periods across data sets, which cannot be disaggregated and compared by race-year.

Unique, new data that correct historical shortcomings related to racial and ethnic classification have recently become available. The contribution of this paper is to summarize recent efforts at refining these data. Such a refinement would aid in minimizing measurement error in the lynching series and the accompanying problems posed in estimation. Opportunities to test a plethora of economic outcomes using lynching data as a regressor have arisen, given these revisions of historical lynching series.

A natural next exercise would be to extend Cook (2011a) and other papers relying at least partly on historical lynching data sets to incorporate many races rather than just two to test the effect of violence on economic activity.<sup>14</sup> Alternatively, the new data might be used to isolate actual whites and to re-estimate the regressions related to economic activity. One could also imagine application of the new data to questions related to the economics of crime and political outcomes, e.g., voting behavior and political competition. Executing any of these ideas using national data sets would constitute future research, as a national lynching data set does not yet exist, as aforementioned.

---

<sup>14</sup> Data on segregation laws may be used in these regressions, as they often named the groups being sanctioned. Nonetheless, other historical economic data may not include Latinos, Native Americans, etc., e.g., wages or schooling.

## References

- \_\_\_\_\_. 2011. "Violence and Economic Activity: Evidence from African American Inventors, 1870–1940." Michigan State University, unpublished manuscript.
- Carden, Art. 2006. *Institutions and Southern Development: Lynching as a Signal of Insecure Property Rights*. Dissertation, Washington University.
- Carrigan, William D. 2006. *The Making of a Lynching Culture: Violence and Vigilantism in Central Texas, 1836-1926*. Urbana-Champaign: University of Illinois Press.
- \_\_\_\_\_ and Clive Webb. 2003. "The Lynching of Persons of Mexican Origin or Descent in the United States, 1848 to 1928." *Journal of Social History* 37(2) (Winter): 411–38.
- Cook, Lisa D. 2004. *National Lynching Data Series*.
- \_\_\_\_\_. 2011a. "Violence and Economic Activity: Evidence from African American Inventors, 1870–1940." Michigan State University, unpublished manuscript.
- \_\_\_\_\_. 2011b. "Converging to a National Lynching Database: Recent Developments and the Way Forward." Forthcoming, *Historical Methods*.
- Cutler, James Elbert. [1905] 1969. *Lynch-law: An Investigation into the History of Lynching in the United States*. New York: Longmans, Green, and Co.
- Fryer, Roland G., and Steven Levitt. 2007. "Hatred and Profits: Getting under the Hood of the Ku Klux Klan." *NBER Working Papers* 13417.
- Gonzalez-Day, Ken. 2006. *Lynchings in the West, 1850–1935*. Durham, NC: Duke University Press.
- Litwack, Leon F. 1979. *Been in the Storm so Long: The Aftermath of Slavery*. New York: Knopf.
- Margo, Robert A. 1984. "Accumulation of Property by Southern Blacks before World War I: Comment." *American Economic Review* 74: 768–76 ().
- NAACP. 1919. *Thirty Years of Lynching in the United States, 1889–1918*. New York: NAACP.
- Pfeifer, Michael J. 2004. *Rough Justice: Lynching and American Society, 1847–1947*. Champaign-Urbana: University of Illinois Press.

Project HAL Data Collection Project.  
<http://people.uncw.edu/hinese/HAL/HAL%20Web%20Page.htm>, last accessed 2004.

Tolnay, S. E., and E. M. Beck. , 1995. *A Festival of Violence: An Analysis of Southern Lynchings, 1882–1930*. Urbana: University of Illinois Press.

“Lynchings: By State and Race, 1882–1968.” Tuskegee University Library and Archives, <http://www.electricprint.com/academic/departement/AandL/AAS/ANNOUNCE/vra/lynch/lynchstats.html>, last accessed May 25, 2011.

Table 1.  
Historical and Contemporary Lynching Data Sets Compared

Data Set	Period	N	Victim Data	Principal Sources
<i>Historical</i>				
<i>Chicago Tribune</i>	1882-1918	3337	R	NW
Tuskegee University	1882-1968	4743	R	CT,NW,TU
NAACP (1919)	1889-1918	3224	R	CT,NW,I
<i>Contemporary, National</i>				
Cook (2004)	1882-1940	4418	R	TU,BT
Carter, et al. (2006a)	1882-1964	4745	R	TU
Pfeifer (2004)	1874-1947	724	N,R, E	NW,BGA,V
Carrigan and Webb (2003)	1848-1928	597	NO [E]	NW,V
<i>Contemporary, Regional and State</i>				
Project HAL (2004)	1882-1930	2806	G,N,R	BT,OTH
Tolnay and Beck (1997)	1882-1930	2805	G,N,R	CT,NAACP,TU, NW
Bailey, Beck, and Tolnay (2011)	1882-1929	900	A,B,G,N,R	BT, CE
Gonzalez-Day (2006)	1850-1935	352	N,NO,R, E	NW, V

Sources: Cook (2011b). The Beck and Tolnay (1997) data are the underlying data for Tolnay and Beck (1995). Other sources are described in the text.

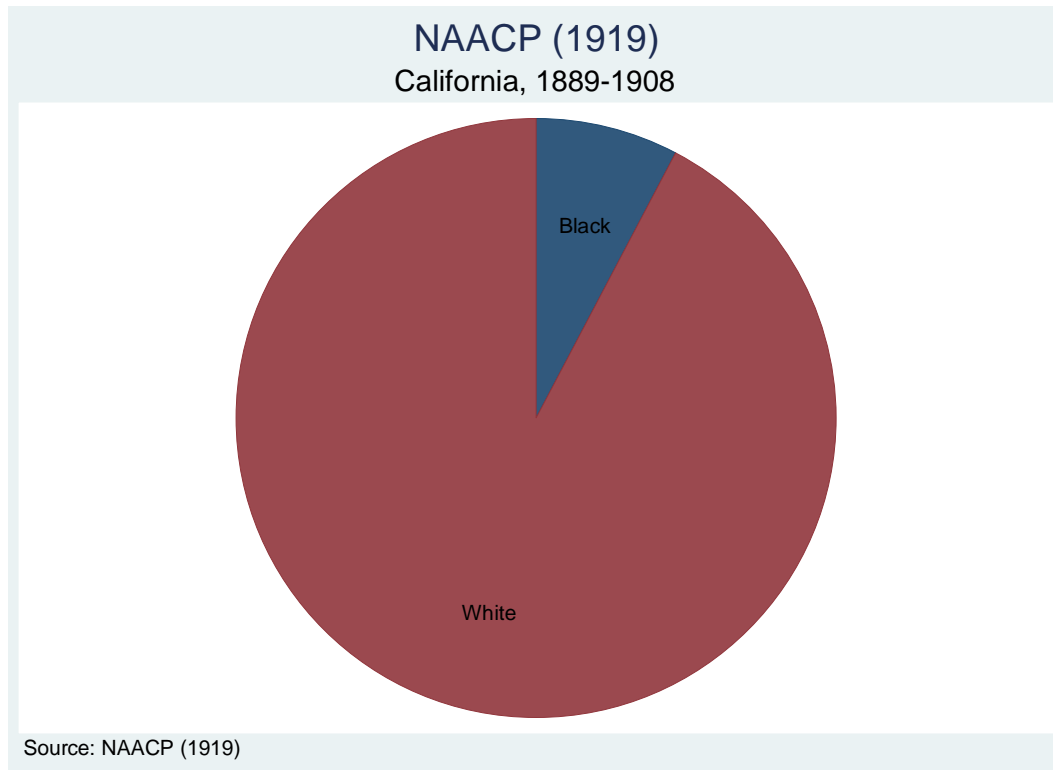
Note: Cutler/Tribune data span 1882 to 1903.

A=age; B=Biographical data; G=gender; N=name; NO=national origin; R=race; E=ethnicity  
I=interviews; NW=newspapers; OTH=other contributors; V=various

BGA=Beck, GA Lynching Project; BT=Beck and Tolnay; CE=Census; CT=*Chicago Tribune*;

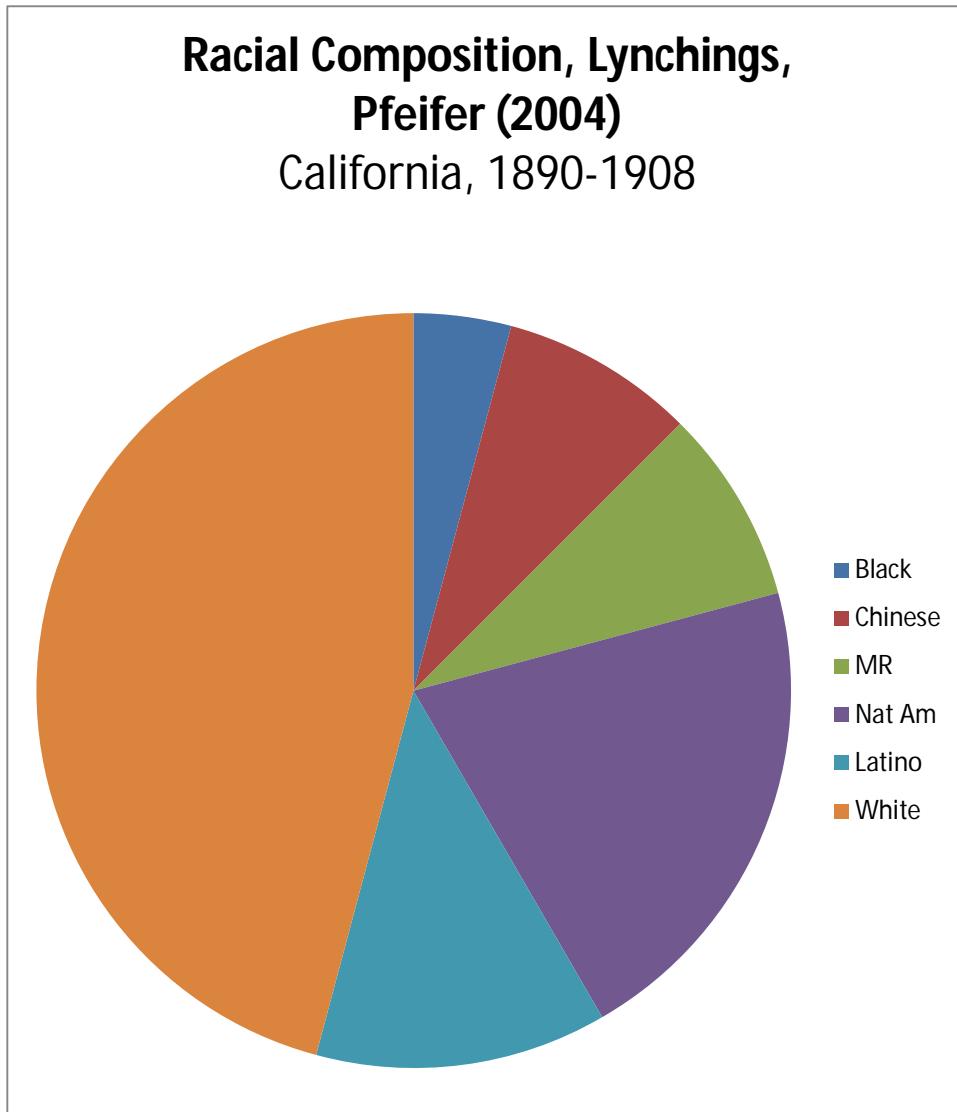
TU=Tuskegee University

Figure 1. Racial Composition, Lynchings



Note:  $N = 26$ .

Figure 2.

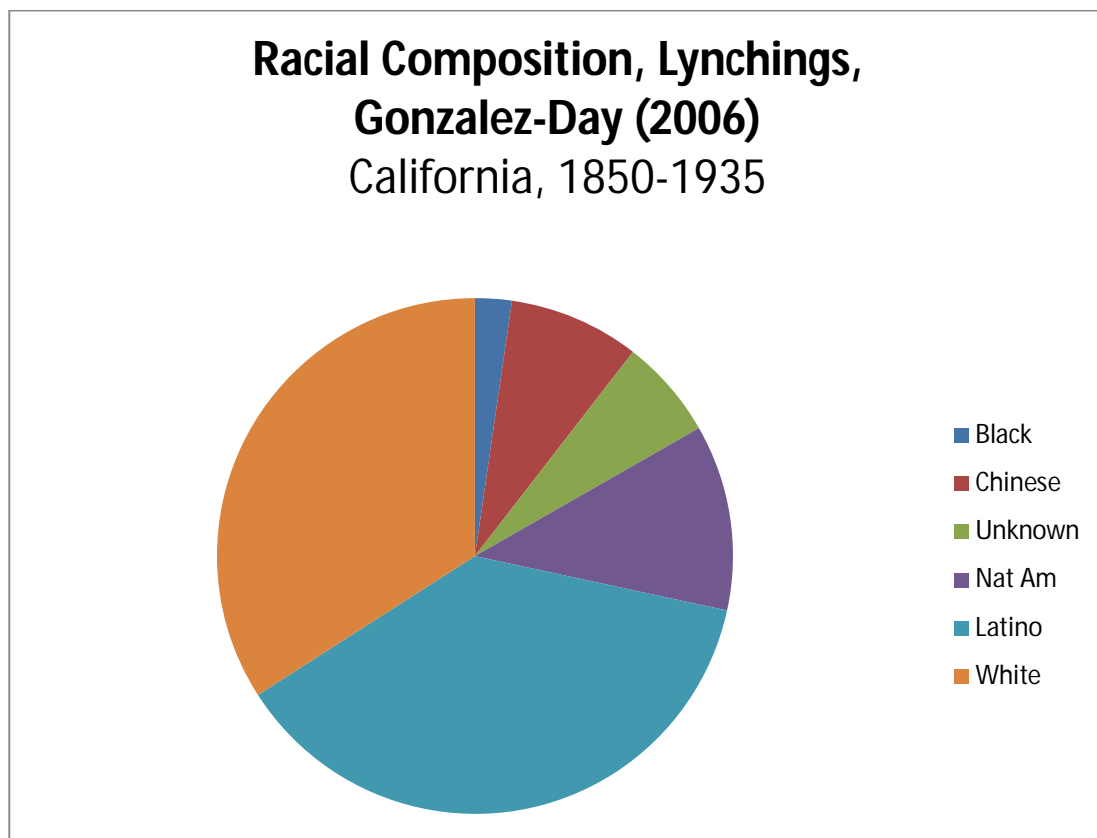


Source: Pfeifer (2004)

Note: MR = mixed race; *N* = 24.



Figure 3.



Source: Gonzalez-Day (2006)

Note:  $N = 352$