Accessing IRS Form 990 Data with Excel

Humberto Barreto

Q.G. Noblitt Professor of Economics and Management

hbarreto@depauw.edu

Michele T. Villinski

Professor of Economics and Management

mvillinski@depauw.edu

DePauw University

Greencastle, IN 46135

Please do not quote without permission.

All comments welcome.

Abstract

Form990.xlsm is a freely available, standalone, macro-enabled Excel workbook that allows users to easily download and analyze IRS Form 990 tax data for nonprofits. This information is extremely difficult to access in machine-readable format, but the IRS recently made the data available as XML files. *Form990.xlsm* directly imports data in separate sheets for each institution. This paper demonstrates how to use this tool by analyzing a small sample of private, selective liberal arts colleges. We compare Form 990 data with IPEDS and find substantial agreement, but important differences. Endowments were different by more than 5% in 7 of 46 observations. Student FTEs had a greater than 5% difference in 22 out 46 schools (several were over 20%). Form 990 data offer researchers an alternative to IPEDS that is arguably of higher quality.

Highlights

- Provides a freely available, standalone, macro-enabled Excel workbook (*Form990.xlsm*) to download and analyze IRS Form 990 data.
- Demonstrates how to use the *Form990.xlsm* tool on a small sample of private, liberal arts schools.
- Compares Form 990 to IPEDS data on a few compensation, tuition, and endowment variables and finds substantial agreement, but important differences.

JEL Codes: C8, I20

Keywords: nonprofit, endowment, tuition, IPEDS

1.0 Introduction

The *Return of Organization Exempt from Income Tax*, known as Form 990, is how nonprofits report financial information to the Internal Revenue Service (IRS). First filed for the 1941 tax year with two pages and three questions (Chasin, Kawecki, & Jones, 2002), Form 990 has become the key source of data on tax-exempt organizations. "For Tax Year 2014, over 293,000 charities exempt under section 501(c)(3) filed Forms 990 with the IRS reporting over $3.7 trillion in assets and nearly $2.0 billion in revenue." (U.S. Department of the Treasury, n.d. b).

On June 16, 2016, "the Internal Revenue Service announced that the publicly available data on electronically filed Forms 990 will now be available for the first time in a machine-readable format through Amazon Web Services (AWS)." (U.S. Department of the Treasury, 2016) Before widespread computer use, Form 990 data were only available by direct request to the IRS and, more recently, only image files have been supplied. Extracting data was tedious and cumbersome. The AWS release is a significant improvement in openness and accessibility because the data are available as XML files, making it much easier to process information from millions of returns. Visit lecy.github.io/Open-Data-for-Nonprofit-Research for more on the campaign to liberate 990 data.

This paper further lowers the costs of access by showing how to use *Form990.xlsm*, a macro-enabled Excel workbook that offers an easy interface to the AWS data. *Form990.xlsm* is freely available, with open-source Visual Basic code, at https://archive.org/details/Form990.xlsm. The workbook contains detailed instructions and documentation.

We used *Form990.xlsm* to download a few variables on a small sample of private, selective liberal arts colleges. We compared the Form 990 data to the most common source used by higher education researchers, the Integrated Postsecondary Education Data System (IPEDS). We find substantial agreement, but there are important differences for individual observations between Form 990 and IPEDS data. Although

not perfect, we believe Form 990 data are superior to IPEDS data and recommend using Form 990 when possible.

The next section reviews Form 990 and IPEDS data sources. Section 3 describes our sample and introduces the *Form990.xlsm* Excel workbook. Our results are in section 4, followed by a conclusion.

2.0 IPEDS and Form 990

Financial data used in research on higher educational institutions comes from three main sources: independent/researcher-generated surveys, IPEDS, and the IRS Form 990. Independent surveys are relatively rare, while IPEDS is the most frequently used resource. The National Center for Education Statistics (NCES), part of the U.S. Department of Education and the Institute of Education Sciences (U.S. Department of Education, n.d. b), collects information on a wide range of financial, student, enrollment, and institutional characteristics from over 7,500 U.S. colleges, universities, vocational, and technical schools each year in the IPEDS reporting process. Every institution that participates in federal student financial aid programs must report to IPEDS (U.S. Department of Education, n.d. a). Data are available to researchers, students, and the public online (U.S. Department of Education, n.d. c). Each institution designates a keyholder, the individual responsible for overseeing the IPEDS data collection and entry process. Typically, the keyholder works with offices across the institution to provide data for Fall (October), Winter (February), and Spring (April) collection deadlines (U.S. Department of Education, 2018). A keyholder described the process and offered the following opinion:

> In the decades since the introduction of IPEDS, many improvements have been made in both upload and download processes. IPEDS uses variance formulae to detect data that are inconsistent with previous years, allowing institutions to verify or correct, and provide an explanation for variation outside the norm. Many collection components have file upload options, which also greatly increase efficiency for keyholders. IPEDS takes great care in verifying the reliability of reported data, following up with keyholders when questions arise. This is the

primary reason that release of data takes longer than researchers and the public like (W. Tobin, personal communication, September 20, 2018).

IPEDS is an omnipresent data source in research on higher education, in part due to the large number of institutions, variables, and years it covers. For example, Hedrick, Wassell, and Henson (2009) relied solely on IPEDS data in their analysis of trends in higher education administrative costs. They examined 17 years of data for 249 schools. In contrast, Powell, Gilleland, and Pearson (2012) combined IPEDS data on institutional characteristics, expenditures, and retention and graduation rates with Carnegie classifications and variables from the National Study of Postsecondary Faculty to devise a model for assessing both efficiency and effectiveness in higher education.

While IPEDS has been the dominant data source, information on financial aspects of higher education and other nonprofit institutions is also available from annual IRS Form 990s. According to the instructions for filing, "Form 990 must be filed by an organization exempt from income tax under section 501(a) … if it has either (1) gross receipts greater than or equal to $200,000 or (2) total assets greater than or equal to $500,000 at the end of the tax year." (U.S. Department of the Treasury, n.d. a). For the 2014 tax year, nearly 300,000 tax-exempt organizations filed Form 990 (U.S. Department of the Treasury, n.d. b), including hospitals, colleges and universities, charitable trusts, and political organizations. Charitable organizations that do not file Form 990 include: some churches and religious-affiliated organizations, certain governmental and political organizations (such as the state or local committee of a political party), and specified organizations with gross receipts below the threshold to file ($50,000). Key categories of data reported on the IRS Form 990 include revenue, expenses, assets, a balance sheet, and names, titles, and compensation of top officers. Researchers should note that, for example, the 2018 Form 990 applies to the fiscal year beginning in 2018; the organization lists the start and end date of the fiscal year in part A of the form. For many higher education institutions, this time period is from July 1, 2017 to June 30, 2018.

Compensation information, though, is based on calendar year. IRS Form 990 data are prepared either by the institution's finance department or by accounting firms hired by the institutions. In our sample, roughly three of four schools used external accountants (35 out of 46 institutions). The Board of Trustees, or a designated subcommittee of the board, reviews the form before submission. In addition, most schools ensure that financial data and compensation information are reviewed by third parties such as auditors and lawyers. As Hyatt (2008, p. 16) notes, "The unfortunate institution whose officials cut corners in completing Form 990 risks an IRS audit, penalties, and public embarrassment."

While the IRS Form 990 dataset is rich, it has been underexploited in the higher education literature. This is likely due to practical barriers to accessing the data. Previous options for gathering data from IRS Form 990 were working directly from pdfs solicited from schools, gathered from websites, retrieved from a service such as Guidestar, or paying someone else to harvest the data. These methods are costly in terms of time, price, or both, and they severely limit flexibility in exploring the data. As recently as 2015, Parker used financial data from IRS Form 990 to assess the costs of switching athletic conferences. He accessed data on a select number of variables for 15 institutions via Guidestar, then entered it into Excel before analyzing it in SPSS. (Parker, 2015)

Another way Form 990 data have been used is through *The Chronicle of Higher Education's* annual compilation of presidential compensation data (e.g., Langbert, 2006). The *Chronicle's* current presidential compensation analysis is an online interactive tool that relies on IRS Form 990 data for private schools (Bauman, Davis, & O'Leary, 2018). The *Chronicle* data are available for $249 per year. However, accessing the IRS Form 990 data directly would yield compensation data for employees beyond the president.

Krishnan, Yetman, and Yetman (2006) compared data from IRS Form 990 with audited financial statements and manual analysis of webpages in their research into fundraising and program expenses in nonprofit organizations. Like Parker, they used Guidestar as a source for IRS Form 990 data. Krishnan et al.'s data

collection process was labor intensive: their IRS Form 990 data were not in spreadsheet form and they obtained some years from Guidestar and others from the National Center for Charitable Statistics database.

While these are merely a few of the studies that have used IRS Form 990 data, they illustrate common issues of accessibility that have existed in the past. Form 990 data have been expensive to gather and cumbersome to analyze. The IRS initiative to make Form 990s available in electronic format through the AWS opens opportunities for automated retrieval, quick importation into spreadsheet and statistical software, and vastly increased flexibility in analysis. In the next section, we present a new, straightforward, customizable tool for accessing and manipulating IRS Form 990 data and give a modest demonstration comparing IRS Form 990 to IPEDS for a small sample of liberal arts colleges. The IRS Form 990 dataset may be better for some purposes than IPEDS. The Form 990 dataset encompasses a wider range of institutions and covers some variables not included in IPEDS. According to Capaldi and Abbey (2011, p. 15), who examined ways universities can better use data to improve their performance, "different universities apply different rules for answering questions on the same [IPEDS] forms." In contrast, Krishnan et al. (2006, p. 399) concluded that, "use of an outside paid accountant increases reliability and usefulness of nonprofit financial reports" as most schools do when completing their tax forms. Differences in the data collection and submission processes for IPEDS and the IRS Form 990 ultimately affect the quality of the financial data available to researchers. One industry insider rated IRS Form 990 as 8 out of 10 on quality of data input on financials, with IPEDS earning 4 out of 10. (B. Kelsheimer, personal communication, October 5, 2017)

3.0 Methods

3.1 Our Sample

We began selecting schools and universities to study by downloading data on all 246 schools in the Carnegie basic classification "Baccalaureate Colleges: Arts and Sciences Focus" (The Carnegie Classification of

Institutions of Higher Education, n.d. a). We used information from Carnegie on school characteristics to narrow the sample to 53 small (or very small) nonprofit, private, 4-year, selective (or more selective), highly residential, exclusively undergraduate colleges and universities focusing on the arts and sciences. Using IRS Form 990 data, we removed seven additional very small schools with fewer than 800 employees, for a final sample of 46 institutions. Table 3.1.1 shows our sample selection procedure.

Sample size after each filter	Filter Description
246	Carnegie basic classification: Baccalaureate colleges: Arts and sciences focus *"Includes institutions where baccalaureate or higher degrees represent at least 50 percent of all degrees but where fewer than 50 master's degrees or 20 doctoral degrees were awarded during the update year."* (The Carnegie Classification of Institutions of Higher Education, n.d. c)
245	Nonprofit
218	Private
81	Undergraduate profile: Arts and sciences focus – no graduate coexistence *"At least 80 percent of bachelor's degree majors were in the arts and sciences, and no graduate degrees were awarded in fields corresponding to undergraduate majors."* (The Carnegie Classification of Institutions of Higher Education, n.d.-e)
63	Graduate program – (not classified) *This indicates the school confers exclusively undergraduate degrees.* (The Carnegie Classificatio of Institutions of Higher Education n.d. b, p. 5)
54	More selective (or selective), lower transfer-in *"Fall enrollment data indicate at least 80 percent of undergraduates are enrolled full-time at these bachelor's or higher degree granting institutions. Test score data for first-year students indicate that these institutions are more selective in admissions (80th to 100th percentile of selectivity among all baccalaureate institutions). Fewer than 20 percent of entering undergraduates are transfer students. Note: 40th to 80th percentile for selective schools.* (The Carnegie Classification of Institutions of Higher Education, n.d. f)
53	Small (or very small), highly residential *"Fall enrollment data indicate FTE enrollment of 1,000 – 2,999 (fewer than 1,000) degree-seeking students at these bachelor's or higher degree granting institutions. At least half of degree-seeking undergraduates live on campus and at least 80 percent attend full time."* (The Carnegie Classification of Institutions of Higher Education, n.d. d)
46	More than 800 employees (Form 990, Part I, Line 5)

Table 3.1.1: Selecting the Sample

The 46 colleges and universities in our sample are distributed across the United States as shown in Figure 3.1.1, which display beginning-of-year endowment.
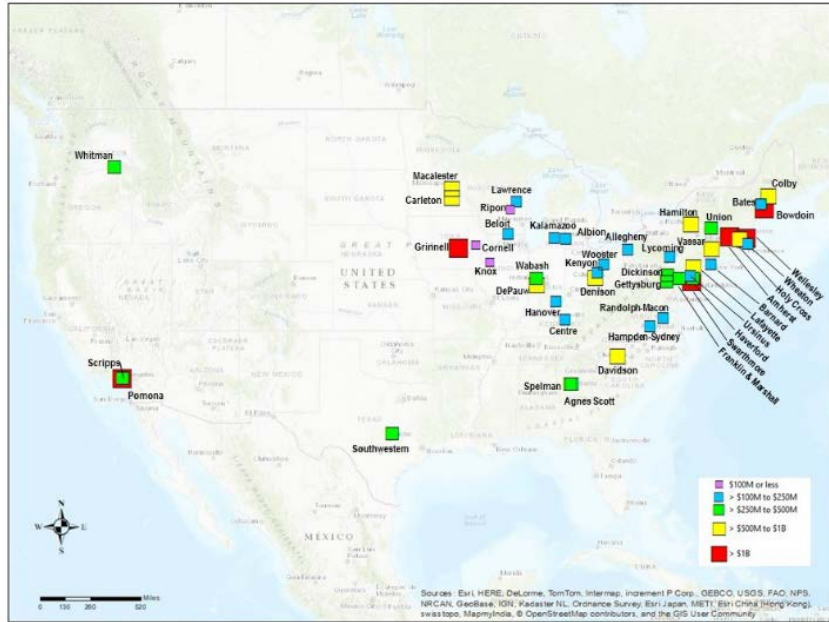

Figure 3.1.2: Endowment in sample schools.

3.2 Form990.xlsm Macro-enabled Excel Workbook

Our macro-enabled Excel workbook is largely self-explanatory, thus, we offer a simple overview here. After downloading the file from https://archive.org/details/Form990.xlsm, open the workbook and make sure that security settings allow macros to run. Review the *Intro* worksheet for details about software requirements and detailed instructions for using the data retrieval tool. Clicking the *EIN* tab takes the user to the sheet that actually accesses the data. The first step is to click the [Get CSV] button and enter the tax year under study. The next step is to enter Employer Identification Numbers (EINs) for each tax filer. Figure 3.2.1 shows an example.

Use the *index* sheet in the workbook (downloaded in step 1) to search by institution name to find the tax filer's EIN. Excel formulas and database functions (e.g., VLOOKUP) can be used in the *index* sheet to return

EINs given institution names. Clicking the [Get XML] button will populate the *Taxpayer Name* column and

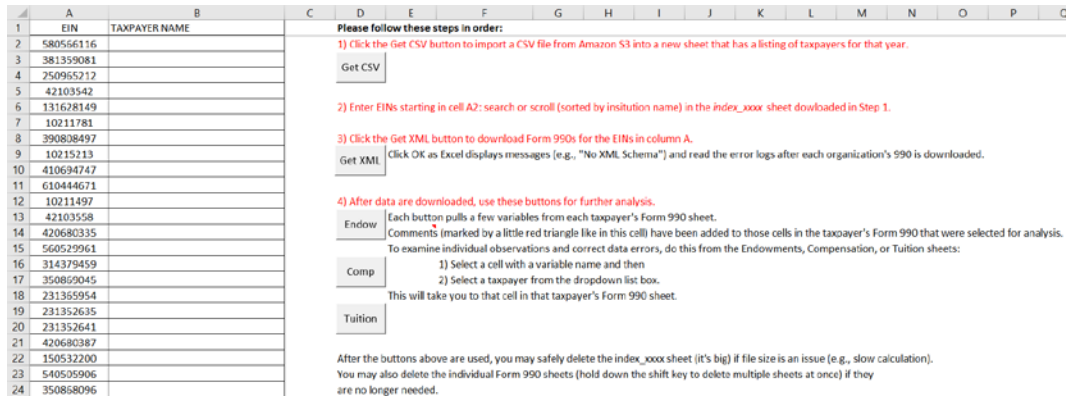download Form 990s as separate sheets in the workbook.

Figure 3.2.1: The *EIN* sheet in *Form990.xlsm*.

Aside from entering or importing the list of EINs, using the tool is a straightforward series of button clicks to:

a) Import tax form index information for a given tax year in CSV format from AWS,

b) enter specific EINs, and

c) download IRS Form 990 data for given EINs into Excel.

While we have focused on a group of small college and universities, these functions of the Form 990 data

retrieval tool can be applied to any institution or group of institutions that filed electronically and whose

form has been included in the AWS releases. Other users might choose to study hospitals, private

foundations, or a different slice of higher education, for instance.

For our higher education data, the Form 990 Excel workbook also includes built-in macros for analyzing

Form 990 after accessing and downloading it from AWS. Three buttons in the *EIN* sheet examine data

related to endowment, tuition, and compensation. Researchers using the Excel workbook to examine

institutions outside higher education, or asking questions different from ours, will branch off in other

directions with full flexibility, once they have retrieved their data.

4.0 Results

Following the instructions above, we used *Form990.xlsm* to download XML data for our sample institutions for the 2013 tax year (July 1, 2013 to June 30, 2014 for most schools). The filing deadline is May 15, 2015, and the 990 forms appear over the next several months in the AWS database.

We then used the endowment, tuition, and compensation buttons to extract a few variables and saved the results in three separate Excel files available at https://archive.org/details/Endowment . We checked the XML data with image (pdf) files to confirm they were the same. Finally, we downloaded these same variables from IPEDS and compared the two sources.

We report summary results below, but all of the data are available in the Excel files. It is easy to identify specific schools and sort by different variables.

4.1 Endowment

The Excel workbook, *Endowment.xlsx*, contains all the data and results described in this section. Form 990 asks schools to report the beginning and ending endowment for a five-year period, as shown in Figure 4.1.1 for one institution in our sample (Barnard College).

Part V **Endowment Funds.** Complete if the organization answered "Yes" to Form 990, Part IV, line 10.					
	(a) Current year	**(b)**Prior year	**b (c)**Two years back	**(d)**Three years back	**(e)**Four years back
1a Beginning of year balance	240,710,000	213,516,000	215,488,726	184,014,662	165,305,949
b Contributions	13,732,348	6,855,381	3,771,498	3,637,098	2,562,445
c Net investment earnings, gains, and losses	39,368,149	31,712,887	4,717,370	37,263,758	25,699,698
d Grants or scholarships	5,741,252	5,366,622	4,968,942	4,393,254	3,297,095
e Other expenditures for facilities and programs	4,842,748	4,584,844	4,179,058	4,488,747	5,584,906
f Administrative expenses	1,178,328	1,422,802	1,313,594	544,791	671,429
g End of year balance	282,048,169	240,710,000	213,516,000	215,488,726	184,014,662

Figure 4.1.1: Example of 2013 Form 990 Endowment data.

We compared the values for beginning of year balance (line 1a, column (a) Current Year) to the IPEDS variable, *Value of endowment assets at the beginning of the fiscal year (F1314_F2_RV)*. We also compared the values for the end of year endowment balance. For Barnard, IPEDS and Form 990 were identical.

We found that 25 schools had exactly or almost exactly the same values in the two data sources, but 7 of 46 schools had a greater than 5% difference in beginning endowment. These same seven schools, with Franklin and Marshall and Amherst having greater than a 10% discrepancy, also had large differences in ending endowment. Figure 4.1.2 displays the results for the full sample.

Figure 4.1.2: Comparing Form 990 and IPEDS Measures of Endowment.

Source: *Comparing!Endowment.xlsx.*

We used Form 990 data to compute compound annual growth rates (CAGR) for the endowment over the five-year period and annual endowment draws. The CAGR includes all contributions to the endowment (fund-raising); investment earnings, gains, and losses; grants or scholarships (drawn from the endowment); other expenditures for facilities and programs; and administrative expenses. For Barnard, the CAGR was

$$CAGR_{Barnard} = \left(\frac{282,048,169}{165,305,949}\right)^{\frac{1}{5}} - 1 \approx 11.3\%.$$

Figure 4.1.3 shows the distribution of CAGRs in our sample. Barnard's 11.3% annualized return was quite

good. The average CAGR for this time period for our 46 schools was about 8.8%. The data show a great deal

of dispersion, with two schools (Lycoming and Knox) growing their endowments at extremely fast rates of

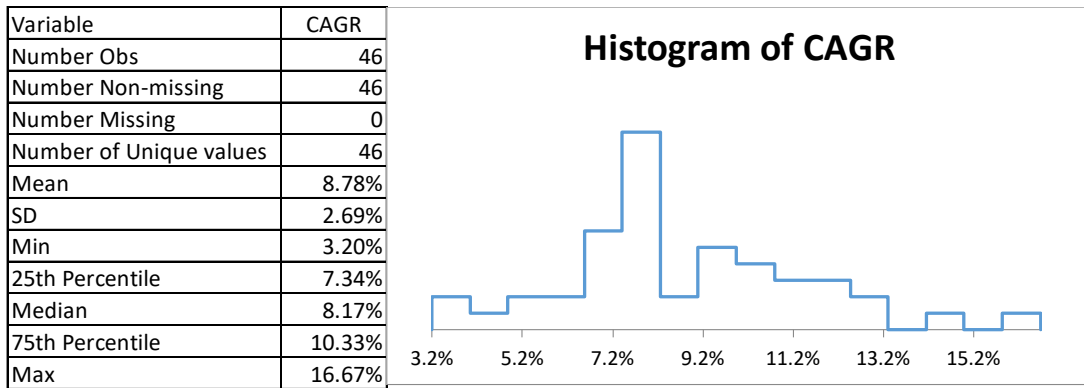14.5% and 16.67%, respectively, while several schools generated CAGRs under 5% per year.

Variable	CAGR
Number Obs	46
Number Non-missing	46
Number Missing	0
Number of Unique values	46
Mean	8.78%
SD	2.69%
Min	3.20%
25th Percentile	7.34%
Median	8.17%
75th Percentile	10.33%
Max	16.67%

Figure 4.1.3: Five-year Endowment CAGRs.

Source: *HistCAGR!Endowment.xlsx.*

Schools also vary widely in their use of endowment draw. Most university administrators want to see the

draw below 5% per year and, at many colleges, presentations of the budget highlight the proposed draw for

the upcoming year. Form 990 data offer a fresh perspective by enabling computation of the realized, or ex

post, draw.

Our annual endowment draw variable computes the grants or scholarships, other expenditures for facilities

and programs, and administrative expenses as a fraction of the average of beginning and ending

endowment amount that year. (See the *Doc* sheet in *Endowment.xlsx* for a demonstration that the average

endowment provides a good approximation when computing the annual draw.) For Barnard, we sum lines d,

e, and f in Figure 4.1.1 and divide by the average endowment balance for each year. For the current year

(the 2013-14 tax year) for Barnard, the draw was

$$Current\ year endowment\ draw_{Barnard} = \frac{5{,}741{,}752 + 4{,}842{,}748 + 1{,}178{,}328}{\frac{240{,}710{,}000 + 282{,}048{,}169}{2}} \approx 4.5\%.$$

Barnard's average draw for the five-year period was 4.9%, which is a little better than the average draw in our sample, as shown in Figure 4.1.4. Lycoming had the lowest average draw by far, at 1.2%, while 8 out of 46 schools had average draws above 6%, with Union College taking the highest average draw of 8.4%.
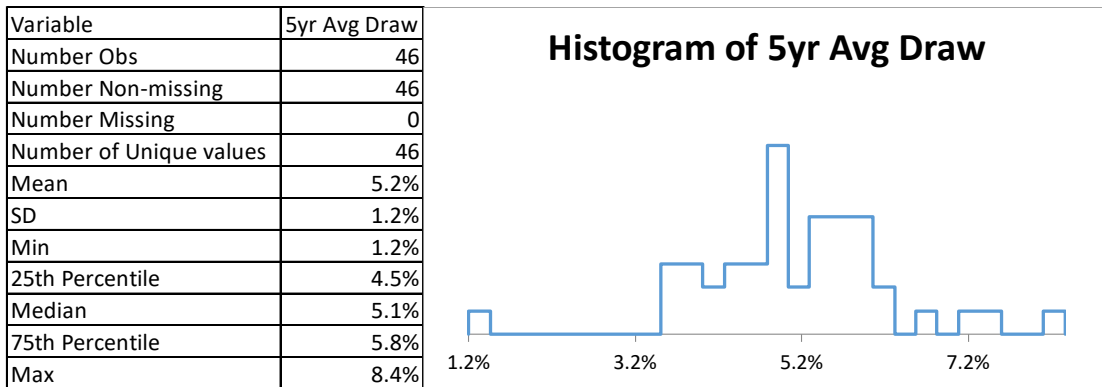
Variable	5yr Avg Draw
Number Obs	46
Number Non-missing	46
Number Missing	0
Number of Unique values	46
Mean	5.2%
SD	1.2%
Min	1.2%
25th Percentile	4.5%
Median	5.1%
75th Percentile	5.8%
Max	8.4%

Figure 4.1.4: Five-year Average Endowment Draws.

Source: *HistDraw!Endowment.xlsx.*

4.2 Tuition and Student Enrollment

The data and results in this section are available in the Excel workbook *Tuition.xlsx*. Schools report the amount of tuition and fees they billed in Part VIII: Statement of Revenue of their Form 990, as shown in Figure 4.2.1 for Dickinson College.

Figure 4.2.1: 2013 Form 990 Tuition and Fees data for Dickinson College.

Like Dickinson, most schools separate out various other revenue sources, but some lump all other sources into a single number. Carleton and Centre colleges are the only two institutions in our sample that combined room and board and tuition. Carleton separated out the room and board charge on their web site and we called Centre to get their room and board charge.

Given billed revenue from tuition and fees for each school, we can compute a high-quality estimate of full-time enrollment (FTE) since total revenue equals price times quantity. Dividing billed tuition and fees by the school's list (or sticker) price for tuition and fees (which is easily available) yields FTE. Figure 4.2.2 compares the Form 990 FTE with the FTE reported in IPEDS. While broadly in agreement, there are substantial differences. For all of the schools in our sample, IPEDS reported an FTE count of 82,626, while the Form 990 estimate was 80,318 FTE students. There were 22 out 46 schools with more than a 5% difference in FTEs, led by Bates, Scripps, and Spelman with differences greater than 20%.
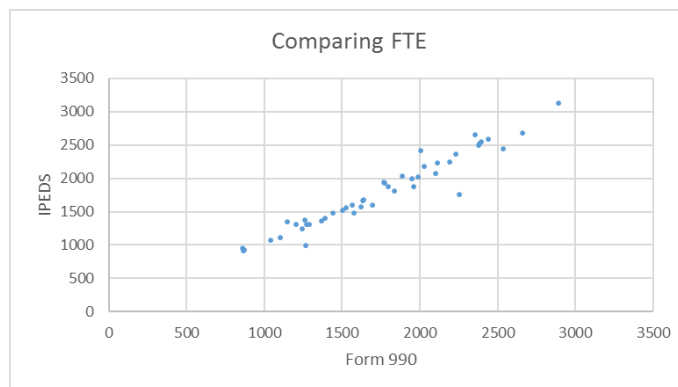

Figure 4.1.2: Comparing Form 990 and IPEDS Measures of FTE.

Source: *Comparing!Tuition.xlsx.*

Along with the endowment draw, administrators at small colleges often report and focus on the financial aid fraction, the proportion of list price tuition that is subsidized by grants from the institution. Since colleges report how much grant aid they give students on Form 990 (line 13 in Part 1, on the first page, reports the

sum of grants in Part IX, column (A), lines 1-3), dividing this by billed tuition tells us the fraction of tuition subsidized by grants to students. Dickinson reported $40,682,351 in grants to students, which is 36.9% of $110,348,110 in gross billed tuition. This is below 46.5%, the average tuition subsidy in our sample of schools. Figure 4.2.3 shows that there is a great deal of variation in how much schools subsidize students, from a low of 27.2% by Scripps to a high of 65.0% by Grinnell.
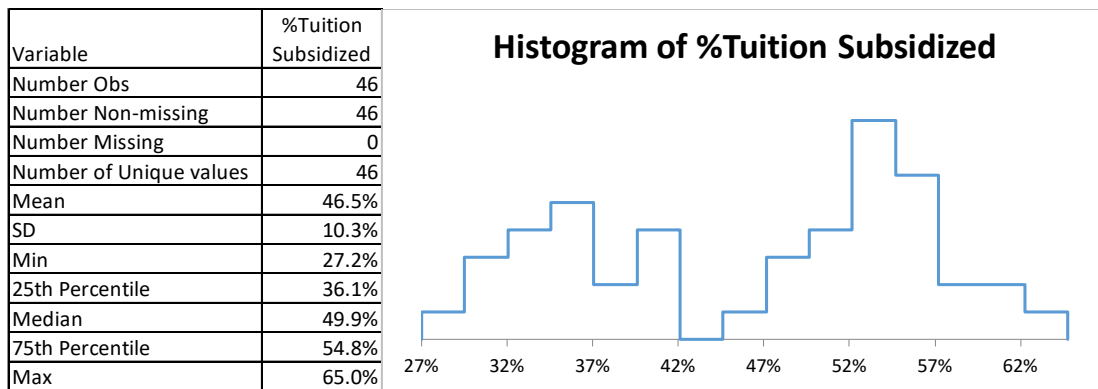
Variable	%Tuition Subsidized
Number Obs	46
Number Non-missing	46
Number Missing	0
Number of Unique values	46
Mean	46.5%
SD	10.3%
Min	27.2%
25th Percentile	36.1%
Median	49.9%
75th Percentile	54.8%
Max	65.0%

Figure 4.2.3: Fraction of Tuition Subsidized by Endowment Grants to Students.

Source: *HistSub!Tuition.xlsx.*

4.3 Compensation

Form 990 requires listing current officers, directors, and trustees, regardless of compensation, and up to 20 current *key employees* (IRS). This is the source of reports on compensation for presidents by the *The Chronicle of Higher Education* and news stories on other highly-paid people, such as coaches and athletic directors. The data for our sample exactly matched the *Chronicle's* data, available behind a pay wall at www.chronicle.com (Bauman et al., 2018).

While IPEDS does not have presidential salary data, we used our small sample to examine the difference in male-female presidential compensation and explore the ratio of presidential to average employee pay. Download *Compensation.xlsx* to access the data and results in this section.

The IRS allows organizations to submit compensation data "for the calendar year ending with or within the organization's tax year." (U.S. Department of the Treasury, n.d. a). Thus, if a new president or other key employee comes on board in July (typical in higher education), Form 990 will show what compensation they received for the months worked in the calendar year. Our sample tax returns are for the 2013-14 fiscal year and organizations are reporting compensation data for the 2013 calendar year.

We removed ten presidents who started in 2013 and, therefore, received partial year compensation for that calendar year and the president of Holy Cross, who does not receive a salary. Figure 4.3.1 shows that there was little difference in total compensation received by male and female presidents in our sample.
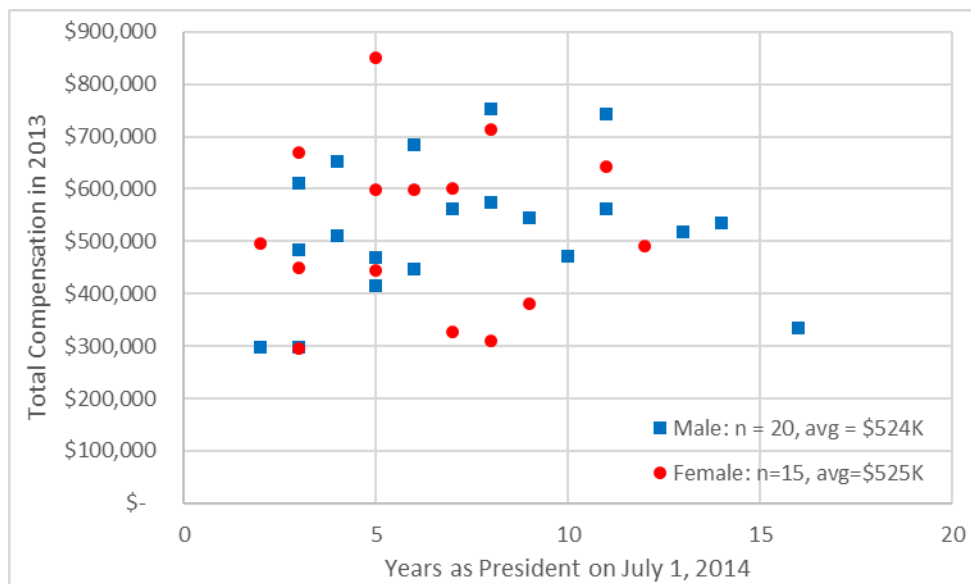

Figure 4.3.1: Total Compensation of Presidents by Gender and Years on the Job.

Source: *CompPresFullYear!Compensation.xlsm.*

We used the information provided in Form 990 data to measure the ratio of the highest paid person to the average pay. Form 990 reports the number of employees that were given a W-2 and the total compensation paid to all employees so calculating the average compensation is simple (see column H in the *Compensation* sheet). We found the maximum total compensation paid to an individual (usually the president, but not always) at each institution and divided by the average compensation (see column I). Figure 4.3.2 shows the wide variation in the high-to-average ratio, with many schools in the teens, but others with the highest paid person receiving 20 and 30 times the average compensation. Bowdoin is the outlier at almost 45. Like presidential pay, this measure is subject to the vagaries of interim leadership and hiring in the middle of the year. Furthermore, some schools include food and other service workers as their employees, while others outsource. Making good comparisons is difficult.

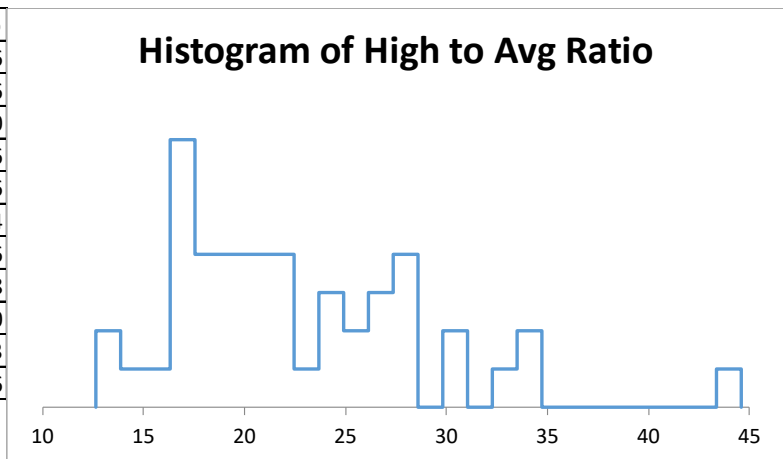| Variable	High to Avg Ratio
Number Obs	46
Number Non-missing	46
Number Missing	0
Number of Unique values	46
Mean	22.6
SD	6.4
Min	12.6
25th Percentile	17.8
Median	21.0
75th Percentile	26.3
Max	44.6

Figure 4.3.2: Highest to Average Compensation Ratio.

Source: *CompPresFullYear!Compensation.xlsm.*

5.0 Conclusions and Future Research

While Form 990s have been available as image (pdf) files for some time, extracting information was tedious and expensive. Thus, the availability of Form 990 data in machine-readable format is a real step forward in transparency and offers a rich new source of information. These data are arguably superior to IPEDS

because Form 990 information is vetted by third parties according to generally accepted accounting principles, submitted to the IRS, and institutions are aware that they could be audited.

Although Form 990 data are not perfect, they are of higher quality than IPEDS, the most common source for higher education research. IPEDS, currently the most common source for higher education research, is a complicated survey managed by a single person (usually in a department of institutional research) coordinating input from various sources (such as admission, finance, and human resources). There are missing data issues and imputation methods are applied. Most researchers download and use IPEDS unquestioningly.

We compared Form 990 data with IPEDS and found important differences. Endowments were different by more than 5% in 7 of 46 observations. We estimated FTEs based on Form 990 billed tuition data and found a greater than 5% difference from IPEDS in 22 out 46 schools, with several schools showing more than 20% differences.

Form 990 data allow for research into areas not covered by IPEDS. For instance, we computed CAGRs as a measure of endowment performance and endowment draws. Both showed substantial variation in our sample. We also found wide dispersion in the fraction of tuition subsidized by financial aid. Finally, Form 990 data offers the opportunity to study individual compensation. We found no gender difference in presidential pay (replicating exactly *The Chronicle of Higher Education's* data which are behind a paywall) and a great deal of variation in high-to-average pay ratios.

There are many ways to access Form 990 XML data, e.g., Python, but *Form990.xlsm* offers a convenient, user-friendly tool to quickly get information in a familiar spreadsheet form. We used it on a few private, selective liberal arts colleges to show how it can be utilized in research, but the universe of nonprofits is much broader than higher education. Form 990 data can be used in research into hospitals, political associations, service organizations, and more.

References

Bauman, D., Davis, T., & O'Leary, B. (2018, July 15). Executive compensation at private and public colleges. *The Chronicle of Higher Education.* Retrieved (2018, August 17) from https://www.chronicle.com/interactives/executive-compensation#id=table_public_2017

Capaldi, E. D., & Abbey, C. W. (2011). Performance and costs in higher education: A proposal for better data. *Change: The Magazine of Higher Learning,* 43(2), 8–15.

Chasin, C., Kawecki, D. & Jones, D. (2002). G. Form 990 [PDF file]. U.S. Internal Revenue Service, Tax Exempt & Government Entities Division. Retrieved (2018, August 17) from https://www.irs.gov/pub/irs-tege/eotopicg02.pdf

Hedrick, D. W., Wassell, C. S., Jr., & Henson, S. E. (2009). Administrative costs in higher education: How fast are they really Growing? *Education Economics,* 17(1), 123–37.

Hyatt, T. K. (2008). New Form 990 will follow your functions. *Trusteeship,* 16(1), 14–18.

Krishnan, R., Yetman, M. H., & Yetman R. J. (2006). Expense misreporting in nonprofit organizations. The Accounting Review, 81(2), 399.

Langbert, M. (2006). How universities pay their presidents. *Academic Questions,* 19(2), 67–81.

Parker, J. P. (2015). *An Analysis of the Financial Impact of Changing Athletic Conference Affiliation within the NCAA* (Doctoral dissertation). Retrieved (2018, August 17) from ProQuest Dissertations and Theses Global. (Accession no. 3745341)

Powell, B. A., Gilleland, D. S., & Pearson, L. C. (2012). Expenditures, efficiency, and effectiveness in U.S. undergraduate higher education: A national benchmark model. *Journal of Higher Education,* 83(1), 102–27.

The Carnegie Classification of Institutions of Higher Education. (n.d. a). About Carnegie classification. Retrieved (2018, August 17) from carnegieclassifications.iu.edu

The Carnegie Classification of Institutions of Higher Education. (n.d. b). Basic classification [PDF file]. Retrieved (2018, August 17) from http://carnegieclassifications.iu.edu/downloads/CCIHE2015-FlowCharts-01Feb16.pdf

The Carnegie Classification of Institutions of Higher Education. (n.d. c). Basic classification description. Retrieved (2018, August 17) from http://carnegieclassifications.iu.edu/classification_descriptions/basic.php

The Carnegie Classification of Institutions of Higher Education. (n.d. d). Size & setting classification description. Retrieved (2018, August 17) from carnegieclassifications.iu.edu/classification_descriptions/size_setting.php

The Carnegie Classification of Institutions of Higher Education. (n.d. e). Undergraduate instructional program classification. Retrieved (2018, August 17) from http://carnegieclassifications.iu.edu/classification_descriptions/ugrad_program.php

The Carnegie Classification of Institutions of Higher Education. (n.d. f). Undergraduate profile classification. Retrieved (2018, August 17) from carnegieclassifications.iu.edu/classification_descriptions/undergraduate_profile.php

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (2018). IPEDS keyholder handbook 2018-2019 [PDF File]. Retrieved (2018, August 17) from http://www.airweb.org/EducationAndEvents/IPEDSTraining/Tutorials/Documents/1819%20ICH%20NKH%20and%20Fall%20Tutorials/IPEDS_NewKeyholder_Handbook_201819.pdf

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (n.d. a). About IPEDS. Retrieved (2018, August 17) from https://nces.ed.gov/ipeds/about-ipeds

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (n.d. b). About us. Retrieved (2018, August 17) from https://nces.ed.gov/about

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (n.d. c). Use the data. Retrieved (2018, August 17) from https://nces.ed.gov/ipeds/use-the-data

U.S. Department of the Treasury. Internal Revenue Service. (2016, June 16). IRS makes electronically filed Form 990 data available in new format IR 2016-87. Retrieved (2018, August 17) from https://www.irs.gov/uac/newsroom/irs-makes-electronically-filed-form-990-data-available-in-new-format

U.S. Department of the Treasury. Internal Revenue Service. (n.d. a). 2017 instructions for Form 990 return of organization exempt from income tax [PDF File]. Retrieved (2018, August 17) from www.irs.gov/pub/irs-pdf/i990.pdf

U.S. Department of the Treasury. Internal Revenue Service. (n.d. b). Charities and other tax-exempt organizations, 2014 [PDF File]. Retrieved (2018, August 17) from www.irs.gov/pub/irs-soi/2014taxexemptorganizationsonesheet.pdf